

8-2015

The intersection of nuclear magnetic resonance and quantum chemistry

Yali Wang

University of Nebraska - Lincoln

Follow this and additional works at: <http://digitalcommons.unl.edu/chemistrydiss>



Part of the [Chemistry Commons](#)

Wang, Yali, "The intersection of nuclear magnetic resonance and quantum chemistry" (2015). *Student Research Projects, Dissertations, and Theses - Chemistry Department*. 60.

<http://digitalcommons.unl.edu/chemistrydiss/60>

This Article is brought to you for free and open access by the Chemistry, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Student Research Projects, Dissertations, and Theses - Chemistry Department by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

THE INTERSECTION OF NUCLEAR MAGNETIC RESONANCE AND QUANTUM
CHEMISTRY

by

Yali Wang

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Chemistry

Under the Supervision of Professor Gerard S. Harbison

Lincoln, Nebraska

August, 2015

THE INTERSECTION OF NUCLEAR MAGNETIC RESONANCE AND QUANTUM CHEMISTRY

Yali Wang, Ph.D.

University of Nebraska, 2015

Adviser: Gerard S. Harbison

Nuclear Magnetic resonance and quantum chemistry have been recognized to be strong tools for probing the structure and dynamics of molecules to further solve chemistry and biological problems. Chemical shift measured by NMR experiment and chemical shielding, molecular energy and molecular structure calculated by quantum chemistry provide extensive information.

Exact analytic gradients, are obtained for cavitation, dispersion and repulsion energies and time-dependent density functional theory for the continuum solvation model, which could be used to probe the structure, dynamics and properties of molecules. Copper in Cu_A azurin is recognized to be coordinated by a structure water molecule by comparing the experimental His120 pK_a reported in literature with quantum mechanical calculation result.

Accurate ¹³C NMR chemical shielding for small organic molecules can be obtained by quantum mechanical calculation by considering electron correlation effect, complete basis set extrapolation and vibrational correction. Basis set incompleteness is found to be the main source of inaccuracy and cannot be removed by applying any fixed correction, but is dependent on the chemical nature of the relevant group. The ¹³C chemical shielding of methyl, ethylene and ethyne carbon is significantly improved by vibrational correction.

Trifluoroacetic acid catalyzed retinoic acid isomerization is recognized to simultaneously decay to polymer by using ^1H NMR method. Common intermediate occurs for the isomerization and all-*trans*, 9-*cis* and 9,13-*dicis* retinoic acid all first convert to 13-*cis* retinoic acid. Free energy changes obtained by NMR experiment compare well with the calculated result using quantum mechanical method done by Professor Harbison.

Solid-State CPMAS NMR method shows that DL-aspartic acid crystalizes to racemic crystals rather than conglomerate over most of its temperature range, which is confirmed by PXRD. In contrast, glutamic acid crystalizes as a conglomerate under normal circumstances.

I dedicate this dissertation to:

Wenzhong Wei, my husband,

Ashley Wei, my daughter,

Alex Wei, my son.

Acknowledgements

I would firstly like to thank my advisor, Professor Gerard Harbison for his excellent guidance and expertise and support. I would also like to thank my committee members Professors Xiao Cheng Zeng, Robert Powers, Mark Wilson, and Hui Li for all their helpful advice. Thanks to all of my lab mates who have helped me greatly, Dr. Xiongjian Wu, Dr. John Persons, Dr. Jun Zhou, Dr. Monica Kinde, Dr. Adrienne Roehrich, Dr. Young-Sik Kye and Seth Black.

To all the people of the NMR Instrumentation Center, past and present, Dr. Joe Dumais, Dr. Martha Morton, Sara Basiaga, Dr. Matt Shortridge and Bradley Worley: thank you for all of your guidance and support. Thank you to all of the UNL Chemistry department staff for all of your hard work and support throughout my time here. I thank the professors for whom I have taught, who taught me how to teach.

I thank my parents, Fulan Zhou and Wenying Wang. Although I never saw them during my graduate study, their love always fills me. They supported me applying for graduate school by help me taking care of my daughter and will help me to take care of my son to support me for my career success. I dedicate this dissertation to my hard working husband and my lovely daughter and son.

Table of Contents

Introduction to NMR and Quantum Mechanical Calculation	1
References	29

Chapter 1

Continuous and Smooth Potential Energy Surface for Cavitation, Dispersion and Repulsion Energies in Polarizable Continuum Model

Abstract	31
1.1 Introduction	32
1.2 Theory	33
1.3 Computational Methodology	35
1.4 Results and Discussion	35
References	38
Tables and Figures	39

Chapter 2

Energy Gradients and Combined Time-Dependent Density Functional Theory and Conductorlike Polarizable Continuum Model Method

Abstract	43
----------	----

2.1 Introduction.....	44
2.2 Theory.....	46
2.3 Implementation and Tests.....	53
2.4 Conclusion.....	60
References.....	61
Tables and Figures.....	63

Chapter 3

Protonation of His120 in CuA Azurin: A Quantum Chemical Study

Abstract.....	70
3.1 Introduction.....	71
3.2 Computational Method.....	72
3.3 Results and Discussion.....	74
3.4 Conclusions.....	79
References.....	80
Tables and Figures.....	82

Chapter 4

Computational Accuracy and Efficiency in NMR Chemical Shielding

Abstract.....	88
4.1 Introduction.....	89

4.2 Computational Details.....	91
4.3 Formaldehyde Gas Phase ^{13}C NMR Experiments.....	94
4.4 Results and Discussion.....	94
References.....	99
Tables and Figures.....	101

Chapter 5

Thermodynamics and Kinetics of Acid-Catalyzed Retinoic Acid Isomerization

Abstract.....	113
5.1 Introduction.....	114
5.2 Materials and Methods.....	115
5.3 Results and Discussion.....	120
5.4 Conclusions.....	124
References.....	126
Tables and Figures.....	127

Chapter 6

Solid-State NMR of Amino Acids, and the Origin of Life

Abstract.....	147
6.1 Introduction.....	148
6.2 Theory.....	153

6.3 Materials and Methods.....	156
6.4 Results and Discussion.....	158
6.5 Conclusions.....	161
Tables and Figures.....	162
References.....	170

List of Tables

Table 1.1 Cavitation, dispersion and repulsion energies	39
Table 1.2 Gradients (au) of the total CPCM/RHF/6-31G* energy.....	40
Table 2.1 Ground state and excited state geometry parameters of PNA.....	63
Table 2.2 Geometrical parameters of acetone and acetone+two water cluster.....	64
Table 2.3 Geometrical parameters of 22-atom and 43-atom model molecule of PYP.....	66
Table 2.4 Relaxation energies of 22-atom and 43-atom model molecules of PYP.....	67
Table 3.1 Optimized geometric parameters for the Cu _A model molecules.....	82
Table 3.2 Computed pK _a of His120 of Cu _A azurin.....	83
Table 3.3 Solvation energies and relative energies between base-form and acid-form model molecules for 1CC3.....	84
Table 4.1 ¹³ C NMR chemical shielding of molecules at the MP2 level.....	102
Table 4.2 ¹³ C NMR chemical shielding of molecules at the B3LYP level.....	103
Table 4.3 ¹³ C NMR chemical shielding of molecules at the HF level.....	104
Table 4.4 ¹³ C NMR chemical shielding RMSD with respect to experimental value.....	105
Table 4.5 ¹³ C NMR chemical shielding offset from linear regression fit.....	106
Table 4.6 ¹³ C NMR chemical shielding of molecules at the MP2 level.....	107
Table 4.7 Vibrational corrections and CCSD(T) electron correlation corrections of ¹³ C NMR chemical shielding.....	108
Table 4.8 Computed ¹³ C NMR chemical shielding of molecules by considering complete basis set effect.....	109
Table 5.1 Chemical shifts and coupling constants of vinyl protons of retinoic acid isomers.....	144
Table 5.2 Rate constants of trifluoroacetic acid catalyzed interconversion of retinoic acid	

in CD ₂ Cl ₂ detected by NMR spectra.....	145
Table 5.3 Equilibrium constants of trifluoroacetic acid catalyzed retinoic acid isomerization.....	146

List of Figures

Figure 1.1 The energy and root mean square gradient (RMSG) profiles.....	41
Figure 2.1 Acetone and acetone+two water cluster optimized in aqueous solution.....	65
Figure 2.2 Optimized 22-atom and 43-atom model molecules of PYP in aqueous solution.....	68
Figure 3.1 144-atom model molecule extracted from X-ray structure 1CC3.....	85
Figure 3.2 Schematic structure of the 144-atom model molecule of the Cu _A center.....	86
Figure 3.3 145-atom acid-form and 144-atom base-form model molecules.....	87
Figure 4.1 Gas phase formaldehyde of ¹ H coupled and decoupled ¹³ C NMR spectra...	101
Figure 4.2 Computed ¹³ C chemical shielding.....	110
Figure 4.3 Fit parameters for the CBS extrapolation.....	111
Figure 4.4 The computed chemical shielding with respect to experimental value.....	112
Figure 5.1 Chemical structure of retinoic acid.....	127
Figure 5.2 Kinetic scheme of TFA catalyzed isomerization of retinoic acid.....	130
Figure 5.3 ¹ H NMR spectrum of vinyl protons of all- <i>trans</i> retinoic acid.....	131
Figure 5.4 ¹ H COSY NMR spectrum of vinyl protons of all- <i>trans</i> retinoic acid.....	132
Figure 5.5 ¹ H NMR spectrum of vinyl protons of 13- <i>cis</i> retinoic acid.....	133
Figure 5.6 ¹ H COSY NMR spectrum of vinyl protons of 13- <i>cis</i> retinoic acid.....	134
Figure 5.7 ¹ H NMR spectrum of vinyl protons of 9- <i>cis</i> retinoic acid.....	135
Figure 5.8 ¹ H COSY NMR spectrum of vinyl protons of 9- <i>cis</i> retinoic acid.....	136
Figure 5.9 ¹ H DQF-COSY NMR spectrum of vinyl protons of I ₂ catalyzed 9- <i>cis</i> retinoic acid.....	137
Figure 5.10 All- <i>trans</i> retinoic acid full spectra.....	138

Figure 5.11 Mole fractions of retinoic acid isomers with respect to time after adding trifluoroacetic acid starting from all- <i>trans</i> retinoic acid.....	141
Figure 5.12 Mole fractions of retinoic acid isomers with respect to time after adding trifluoroacetic acid starting from 13- <i>cis</i> retinoic acid.....	142
Figure 5.13 Mole fractions of retinoic acid isomers with respect to time after adding trifluoroacetic acid starting from 9- <i>cis</i> retinoic acid.....	143
Figure 6.1 Theoretical phase diagrams for mixture of L and D enantiomers.....	162
Figure 6.2 Solubilities of L-aspartic acid and DL-aspartic acid.....	163
Figure 6.3 The fit solubilities of L-aspartic acid.....	164
Figure 6.4 Phase diagram for crystallization of mixtures of L and D aspartic acid.....	165
Figure 6.5 CP-MAS ^{13}C NMR spectra of aspartic acid.....	166
Figure 6.6 PXRD spectra of aspartic acid.....	167
Figure 6.7 CP-MAS ^{13}C NMR spectra of aspartic acid evaporated at 91 °C.....	168
Figure 6.8 CP-MAS ^{13}C NMR spectra of glutamic acid.....	169

Introduction to NMR and Quantum Mechanical Calculation

Abstract

The chemical shielding is the most familiar and universal parameter in the field of nuclear magnetic resonance (NMR). It may be calculated or measured experimentally; it may be studied by gas phase, liquid phase, or solid phase NMR, as well as by quantum mechanical methods. This chapter will introduce NMR both from a theoretical and an experimental standpoint, and quantum theory. The following three chapters will use quantum chemistry to briefly discuss cavitation, dispersion and repulsion energies in a polarizable continuum model, time-dependent density functional theory in a polarizable continuum model, and histidine pK_a calculation for Cu_A. The last three chapters of this dissertation will discuss NMR chemical shielding and shifts: gas phase ^{13}C NMR experiments and quantum mechanical calculation for small molecules, ^{13}C solid phase NMR experiments for defining the crystalline states of aspartic acid, and liquid phase ^1H NMR experiments on retinoic acid isomerization.

Nuclear Magnetic Resonance Theory

Nowadays, X-Ray diffraction, Nuclear Magnetic Resonance (NMR) and Mass Spectroscopy are three major techniques for molecular structure determination. NMR spectroscopy is one of the most powerful methods to study the structure and dynamics of molecules, and the only that can be used in the gas phase, liquid phase and solid phase, spanning the range from single atoms to biological macromolecules.

NMR History

The first NMR experiment was conducted early in 1937 by Rabi by using a molecular beam^[1], and the first studies of NMR in bulk materials were carried out, respectively, by the Purcell group on paraffin^[2] and by the Bloch group on water^[3]. Since then, NMR has grown from a physical curiosity to an indispensable technique in a wide variety of fields.

Chemists were attracted by the discovery that the precise NMR frequency of a nucleus depends on the state of its chemical environment; it was discovered in 1950 that two nitrogen nuclei^[4] and two fluorine nuclei^[5] had different resonance frequencies in the same molecule. NMR began to be used to probe the structure of molecules when separate resonance lines for chemically different protons in the same molecule was detected in 1951^[6] and the existence of the chemical shift was appreciated.

Most modern NMR studies depend on the application of sequences of radio frequency (*rf*) pulses. Fourier-transform NMR is usually, if unfairly credited to Richard Ernst and Weston Anderson in 1966^{[7][8]}. Jean Jeener displayed the results of using pulse sequence in terms of two separate frequency scales in 1971^[9], which was developed into two-dimensional NMR.

NMR theory

Three quantities can be measured in a high resolution NMR spectrum: frequencies of the resonance lines, their intensities, and width or shapes of the lines.

An atomic nucleus has spin, which gives it an angular momentum (p) and a magnetic moment (μ). When nuclei experience a magnetic field B , the nuclear magnetic moments interact with the field to give quantized energy levels separated by energy ΔE with the orientation of the spin differing between the two levels.

$$\Delta E = \pm \left(\frac{\mu}{p} \right) B$$

Energy can be resonantly absorbed by the nuclear spin system, permitting nuclear magnetic resonance.

In solids, intermolecular interactions broaden resonance lines due to nuclear spin interactions. This played an important role in developing NMR theory and in elucidating structure of solids. In liquids and gases, rapid molecular tumbling causes many nuclear spin interactions to average to zero, which narrows the peaks. By using magic angle spinning (MAS, sample spinning at angle $\theta_m = 54^\circ 44' 8'' 12'''$ ($\tan \theta_m = \sqrt{2}$) with respect to the static magnetic field), solid molecules can also give high resolution NMR spectra.

From a quantum mechanical viewpoint, the nuclear spins interact with each other and with external magnetic fields, therefore, the nuclear Hamiltonian comprises external terms due to the interaction between nuclear spins and the external magnetic field, and internal terms due to internal interactions between nuclear spins. These internal interactions include chemical shielding, dipole-dipole, J-coupling, spin-rotation and quadrupolar couplings.

In normal NMR experiments, external interactions are much larger than internal interactions. The Hamiltonian of a diamagnetic system can be written as^[10]:

$$H = H_{ext} + H_{int} = H_{0I} + H_{1I} + H_{int}$$

$$= (H_{0I} + H_{1I}) + (H_{CS} + H_D + H_J + H_{SR} + H_Q)$$

The two terms in the first bracket, called the Zeeman interaction, are due to the interaction between spin and external magnetic field. Hamiltonian H_{0I} is the coupling of spin I to the static longitudinal external field B_0 , and H_{1I} , the coupling of spin I to the transverse magnetic field B_1 generated by the radiofrequency coil. These two Hamiltonian can be expressed as:

$$H_{0I} = -\gamma_I \mathbf{I} \cdot \mathbf{B}_0$$

$$H_{1I} = -\gamma_I \mathbf{I} \cdot \mathbf{B}_1$$

Where

$$\mathbf{I} = (I_x, I_y, I_z)$$

$$\mathbf{B}_0 = (B_x, B_y, B_z)$$

$$\mathbf{B}_1(t) = 2(B_{1x}(t), B_{1y}(t), B_{1z}(t)) \cos \omega t$$

γ_I is the gyromagnetic ratio of the nucleus I .

In general, the static longitudinal field B_0 is defined to be along the z direction, so $H_{0I} = H_Z = -\gamma_I B_0 I_Z$ and $-\gamma_I B_0$ is defined as the Larmor frequency of spin I . The magnitude of the transverse magnetic field oscillates around the spectrometer reference frequency, which is nonzero only during *rf* pulse.

The terms in the second bracket are due to internal interactions. The internal spin interaction contain information about the local structure and are usually treated as

perturbations. In this current dissertation, only high resolution solid, liquid and gas NMR were studied, and therefore the internal interaction terms will be briefly discussed.

Chemical shielding

A nucleus experiences both the applied external magnetic field and the screening or shielding field of the electrons around the nucleus. Thus the magnetic field \mathbf{B} at the nucleus can be expressed as

$$\mathbf{B} = \mathbf{B}_0 - \sigma \mathbf{B}_0 = \mathbf{B}_0 (1 - \sigma)$$

where σ is the second-rank Cartesian tensor called the chemical shielding tensor.

The interaction of a nuclear magnetic moment with an external field \mathbf{B} is given by

$$E = -\mu_A \cdot \mathbf{B}$$

where \mathbf{B} is an external static homogeneous magnetic field, and μ_A is the spin magnetic moment of the nucleus. Experimentally, the external field is shielded by the electrons outside of the molecules and atoms, and the interaction energy can be expressed as

$$E = -\mu_A \cdot (1 - \sigma) \cdot B$$

The chemical shielding tensor is calculated as a second derivative of the molecular energy with respect to the external magnetic field and nuclear magnetic moment, and given by

$$\sigma_A = \left. \frac{\partial^2 E}{\partial B \partial \mu_A} \right|_{B=0, \mu_A=0}$$

The chemical shielding describes the interaction between nucleus and the magnetic field produced by the electrons around the nucleus. The chemical shielding Hamiltonian (H_{CS})

of the spin I is the coupling of the spin I with the magnetic field \mathbf{B}_0 through the shielding tensor σ .

$$H_{CS} = \gamma_I \mathbf{I} \cdot \sigma \cdot \mathbf{B}_0$$

σ can be expressed as:

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix}$$

Since this coupling is established via the electronic surroundings, the investigation of shielding interactions can give valuable information about the electronic state of molecules.

In this dissertation, only the symmetric part of the shielding tensor will be considered.

The symmetric tensor σ_s can be expressed as following:

$$\sigma_s = \begin{pmatrix} \sigma_{xx} & \frac{\sigma_{xy} + \sigma_{yx}}{2} & \frac{\sigma_{xz} + \sigma_{zx}}{2} \\ \frac{\sigma_{xy} + \sigma_{yx}}{2} & \sigma_{yy} & \frac{\sigma_{zy} + \sigma_{yz}}{2} \\ \frac{\sigma_{xz} + \sigma_{zx}}{2} & \frac{\sigma_{zy} + \sigma_{yz}}{2} & \sigma_{zz} \end{pmatrix}$$

It can be diagonalized to give three principal values.

$$\sigma = \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix}$$

The isotropic shielding, a primary parameter in solid, liquid and gas phases, can be expressed as:

$$\sigma_i = \frac{\sigma_{11} + \sigma_{22} + \sigma_{33}}{3}$$

The principal values of chemical shielding may be reparameterized in terms of three values: the isotropic shielding (σ_i), the chemical shielding anisotropy ($\Delta\sigma$) and the asymmetry parameter (η). These are often used in the solid state NMR literature.

$$\Delta\sigma = \sigma_{33} - \sigma_i - \frac{\sigma_{11} - \sigma_i + \sigma_{22} - \sigma_i}{2} = \sigma_{33} - \frac{\sigma_{11} + \sigma_{22}}{2} = \frac{3}{2}(\sigma_{33} - \sigma_i)$$

$$\eta = \frac{\sigma_{22} - \sigma_{11}}{\sigma_{33} - \sigma_i}$$

In an isotropic liquid or gas, the anisotropic chemical shift Hamiltonian disappears after applying secular truncation and motional averaging. The chemical shielding Hamiltonian then becomes:

$$H_{CS}^{iso} = \gamma_I \sigma_{iso} B_0 I_Z$$

if the field, as is conventional, is along z direction. The chemically shifted resonance frequency ω , called Larmor frequency, is given by:

$$\omega = -\gamma_I B_0 I_Z (1 - \sigma_{iso})$$

In NMR spectroscopy an experimental parameter called the chemical shift δ is the change of nuclear shielding of target nucleus with respect to a reference nucleus. Chemical shifts are measured and reported relative to some reference in the dimensionless unit of parts per million (ppm), which is independent of the rf frequency or magnetic strength. Chemical shifts can be expressed as

$$\begin{aligned} \delta &\equiv \frac{\omega_{sample} - \omega_{reference}}{\omega_{reference}} \times 10^6 = \frac{(1 - \sigma_{sample}) - (1 - \sigma_{reference})}{1 - \sigma_{reference}} \times 10^6 \\ &= \frac{\sigma_{reference} - \sigma_{sample}}{1 - \sigma_{reference}} \times 10^6 \approx (\sigma_{reference} - \sigma_{sample}) \times 10^6 \end{aligned}$$

Dipole-dipole Coupling

A nuclear spin behaves as a magnetic dipole, which produces a local field and adds or subtracts from the external magnetic field according to the orientations of the spins relative to the external magnetic field. This type of nuclear interaction is called dipole-dipolar coupling. In solids, the dipolar coupling can cause significant line broadening due to its orientation dependence and can be used to determine internuclear distances and obtain structural information about molecules. For molecules in rapid and random motion as in liquid and gas, the dipole-dipole interaction averages to zero. Magic angle spinning in solid can also remove dipole-dipole coupling and improve resolution. In the current dissertation it is not necessary to consider dipolar interactions, so this interaction and its Hamiltonian H_D will not be discussed further.

J-coupling

The J -coupling is an electron coupled spin-spin interaction, which is mediated by electrons through chemical bonds. Ramsey and Purcell initially recognized that it involves electrons that form chemical bonds^[11]. The J -coupling splits lines into multiplets. J -coupling is usually smaller in magnitude than the chemical shift. J couplings can be positive and negative. The absolute sign of J the coupling can be determined from double resonance experiments and can not be found from ordinary high resolution NMR spectra. Most detected J -couplings are through three bonds. Coupling through more than three bonds is called long-range coupling and is of considerable interest in stereochemistry.

The J coupling Hamiltonian can be expressed as

$$H_J = \sum_{i \neq j} \mathbf{I}_i \cdot \mathbf{J} \cdot \mathbf{I}_j$$

where \mathbf{J} is

$$\mathbf{J} = \begin{pmatrix} J_{xx} & J_{xy} & J_{xz} \\ J_{yx} & J_{yy} & J_{yz} \\ J_{zx} & J_{zy} & J_{zz} \end{pmatrix}$$

The \mathbf{J} -coupling tensor also depends on molecular orientation. The \mathbf{J} -coupling interaction is often ignored in solid state NMR since it is much smaller than the dipole dipole interaction, quadrupole coupling and chemical shift interaction. In liquid state, molecules move randomly and fast, therefore, the directional dependence will be averaged out. The \mathbf{J} -coupling tensor becomes the J coupling constant and can be expressed as

$$J = \frac{J_{xx} + J_{yy} + J_{zz}}{3}$$

Spin-rotation interaction

The nuclear magnetic moments may interact also with the magnetic field induced by the molecular rotation. This is called the spin-rotation interaction. The spin-rotation interaction is due to the breakdown of the Born-Oppenheimer approximation (it assumes nucleus does not move and only electrons move), which suppose nucleus does not move, and it is a nonadiabatic effect^[12]. Spin-rotation constants can be accurately determined from measurements of NMR relaxation times, microwave experiments^[13] and in molecular beam experiments^[14].

The spin-rotation interaction is usually only considered in liquids and gases since the molecular motion and rotation are restricted in solids. The isotropic average of spin-rotation Hamiltonian can be expressed as

$$H_{SR} = c\mathbf{I} \cdot \mathbf{J}$$

where c is the spin-rotation constant.

Quadrupole coupling interaction

Nuclei having a spin greater than $\frac{1}{2}$ possess a quadrupole moment and have a quadrupole coupling interaction. The quadrupole coupling interaction and its Hamiltonian (H_Q) will not be further discussed since current research only considers spin $\frac{1}{2}$ nuclei ^1H and ^{13}C .

In summary, the interaction Hamiltonian in a liquid and gas sample is represented by isotropic chemical shift, scalar spin-spin interaction and spin-rotation interaction. All anisotropic interactions such as chemical shift anisotropy, dipole-dipole interaction, quadrupolar interaction are averaged to zero due to the rapid isotropic molecular motion.

The Hamiltonian in a liquid may be expressed in the isotropic average as:

$$H = -\gamma_I B_0 I_Z (1 - \sigma_I) - \gamma_S B_0 S_Z (1 - \sigma_S) + \mathbf{J} \cdot \mathbf{S} + c \mathbf{I} \cdot \mathbf{S}$$

In a solid sample, all the interactions are retained and may be used to monitor the symmetry properties and electronic state of the solid. In the current work, only cross-polarization magic angle spinning (CPMAS) NMR, combining cross polarization and magic angle spinning, a technique introduced initially by Pines^[15] then refined by Schafer^[16], was used for powdered solid samples.

Magic angle spinning (MAS) is used to average all anisotropies, leaving only the isotropic part of the tensor interactions. The natural line broadening due to dipolar and quadrupolar interactions masks other weak interactions such as shift interactions and scalar coupling, which can be overcome by MAS, sample spinning at angle $\theta_m = 54^\circ 44' 8'' 12'''$ ($\tan \theta_m = \sqrt{2}$) with respect to the static magnetic field. The rapid rotation of solid samples narrows their NMR spectra by making the average dipolar

interaction vanish, although this is usually supplemented or replaced by high-power dipolar decoupling. Shielding anisotropies are also averaged to zero and only the isotropic shift is retained.

In order to obtain a high signal to noise ratio, rare spins such as ^{13}C in current work are first polarized by spin order transfer from the abundant spin. This is called cross-polarization (CP). Cross polarization transfers polarization from high abundance spin to low abundance spin by using the magnetic dipole-dipole couplings. If the Hartmann-Hahn matching condition $\gamma_I B_I = \gamma_S B_S$ applies, the I -spin system can relax towards thermal equilibrium by transferring some of its polarization to the S -spin system. This cross-relaxation pathway uses the exchange term in the dipolar Hamiltonian and hence requires spatial proximity and a reasonable amount of rigidity on the NMR timescale. CPMAS NMR is determined by a Hamiltonian similar to which determines high-resolution spectra of liquid and gas. The Larmor frequency of a given spin in CPMAS can be

$$\omega_S = -\omega_o \left[\sigma_{iso} + A_1 \cos(\omega_R t + \gamma) + B_1 \sin(\omega_R t + \gamma) + A_2 \cos(2\omega_R t + 2\gamma) + B_2 \sin(2\omega_R t + 2\gamma) \right]$$

$$A_1 = \frac{2}{3} \sqrt{2} \sin \beta \cos \beta \left[\cos^2 \alpha (\sigma_{11} - \sigma_{33}) + \sin^2 \alpha (\sigma_{22} - \sigma_{33}) \right]$$

$$B_1 = \frac{2}{3} \sqrt{2} \sin \beta \sin \alpha \cos \alpha (\sigma_{11} - \sigma_{22})$$

$$A_2 = \frac{1}{3} \left[(\cos^2 \beta \cos^2 \alpha - \sin^2 \alpha) (\sigma_{11} - \sigma_{33}) + (\cos^2 \beta \sin^2 \alpha - \cos^2 \alpha) (\sigma_{22} - \sigma_{33}) \right]$$

$$B_2 = -\frac{2}{3} \cos \beta \sin \alpha \cos \alpha (\sigma_{11} - \sigma_{22})$$

α , β and γ are Euler angles corresponding to a rotation from principal axis frame to the rotor axis frame, and ω_R is the sample spinning frequency. In the above ω_S equation, the first term is the isotropic term, and other parts are frequencies oscillating at ω_R and $2\omega_R$

about the isotropic frequency. If $\sigma_S > \Delta\sigma$, these oscillations are negligible. However, when $\sigma_S < \Delta\sigma$, sidebands are present, which can be used to obtain the chemical shift anisotropy and asymmetry parameter by analyzing the intensities of the sidebands.

Experiment

The static magnetic field B_0 is generated by the solenoid coil of the superconducting magnet.

A short, intense rf pulse, produced by transmitter coil located in the probe, is used to excite sample nuclei and permit the observation of nuclear resonance. Probes have been designed for liquids and gases, or high rf power and sample spinning of solids, or for the study of samples at high pressure and temperatures. Fourier transformation can be used to disentangle the many frequencies that are present in the free induction decay, which is the time response to the pulse.

In liquid phase NMR, a homogeneous solution in a glass tube was used. Solution NMR shifts are determined with respect to an internal reference such as tetramethylsilane (TMS) or by using a deuterium based field frequency lock. A high resolution spectrum is obtained by using homogeneous field, supported by shimming and locking.

In gas phase NMR, a gas sample is sealed into special tube called a J-Y tube. Gas phase shifts are determined with respect to an internal reference such as H_2 or CO. Very short relaxation times help to obtain good spectra within several hours for gas samples, even though the sample concentration is low.

Powder samples were used in the current solid state NMR work. Powder samples are made up of randomly orientated crystals. Since nuclear spin interactions depend on the

crystal orientation, the NMR spectrum of a powder sample gives powder patterns or broad lines due to different spectral frequencies produced by different orientations, which will obscure nuclear interactions for bigger molecules with too much spectral overlap. Magic-angle spinning offers the possibility of high resolution spectra. The spinning rate should be comparable to the width of the spectrum of a static sample. Solid state ^{13}C magic angle spinning (MAS) NMR spectra are referenced externally using a probe which does not incorporate a field frequency lock. Admantane is an external ^{13}C standard for solids NMR, and has been reported with better than ± 0.03 ppm accuracy. CPMAS is now a routine method to investigate organic and inorganic solids.

Quantum mechanical calculations

With the recent rapid development of both computer hardware and software packages, modern *ab initio* calculations have become a major theoretical approach for studying electronic structure and properties of molecules. In principle, the Schrödinger equation is solved numerically to get electronic wave functions, which are utilized to obtain molecular properties by evaluating the expectation values of corresponding operators.

The Schrödinger equation includes differentiation with respect to time (t) and position (r) and can be expressed as:

$$\mathbf{H}(r,t)\Psi(r,t) = i \frac{\partial \Psi(r,t)}{\partial t}$$

where \mathbf{H} is the Hamiltonian operator, Ψ is the wave function.

In atomic units, the Hamiltonian for N electrons and M nuclei is

$$\mathbf{H} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}}$$

where M_A is the ratio of the mass of nucleus A to the mass of electron, Z_A is the atomic number of nucleus A , and the Laplacian operators involve differentiation with respect to the coordinates of the i th electron and the A th nucleus. The first term and second term are, respectively, the kinetic energy of electrons and nucleus; the third term is the coulomb attraction of electrons and nuclei; the last two terms are the repulsion between electrons and between nuclei, respectively.

The Born-Oppenheimer approximation is central to quantum chemistry. It states that the electrons in a molecule can be approximately treated as moving in the field of fixed nuclei since nuclei move more slowly due to their larger masses. Hence, the Hamiltonian can be simplified to contain only three terms: kinetic energy of electrons, coulomb attraction energy of electrons and nuclei, and repulsion energy of electrons.

Finding and describing approximate solutions to the electronic Schrödinger equation has been the major task of quantum chemists. Hartree-Fock (HF) self-consistent-field (SCF) approximation, the procedure for solving the Hartree-Fock equation is called the self-consistent-field (SCF) method, is a starting point for more accurate approximations such as Møller-Plesset second order perturbation theory (MP2) and coupled cluster theory (CC), both of which include electron correlation effects.

Ab initio quantum chemistry methods in quantum chemistry, computational chemistry based on quantum mechanical theory, are widely used to calculate the magnetic properties of a molecular electronic system perturbed by externally applied electromagnetic fields and nuclear magnetic and electric moments. A number of methods have been developed for the calculation of molecular second-order magnetic response properties. Several benchmark calculations of NMR parameters for small molecules have

been obtained by using highly accurate methods and large basis sets (a set of functions) and have been compared with experiment. The computation of larger molecular systems cannot use these methods because of calculational expense. In the literature, various approaches have been used to study basis set convergence of nuclear shielding.

Chemical shielding is a major NMR property that has been widely studied by different *ab initio* methods such as HF, MP2, coupled cluster theory with single, double and triple excitation (CCSDT) and density functional theory (DFT). In this thesis, GAMESS^[17], Gaussian^[18] and Cfour^[19] quantum mechanical software packages are the working software for chemical shift and shielding calculations.

Hartree-Fock

Hartree-Fock (HF) is the basis for other quantum mechanical calculations. Hartree proposed the formalism of a self-consistent field^[20], Fock expanded it and Roothaan described the matrix algebra equations for computation^[21]. The essence of the Hartree-Fock approximation is to replace complicated many-electron problems by the one electron problem in which electron-electron repulsion is treated in an average way. Hartree first uses antisymmetrized spin-orbitals as the wave function. The Hartree-Fock equation is an eigenvalue equation of the form

$$f(i)\chi_a(i) = \epsilon_a\chi_a(i)$$

The orbital energy of the spin orbital $\chi_a(i)$ is $\epsilon_a(i)$. $f(i)$ is an effective one-electron operator, and can be expressed as

$$f(i) = h(i) + v^{HF}(i)$$

$h(i)$ called core-Hamiltonian operator is the kinetic energy and potential energy for the i electron attraction to nuclei, and expressed as

$$h(i) = -\frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{r_{iA}}$$

and the effective one-electron potential operator $v^{HF}(i)$, called the Hartree-Fock potential, is the average potential experienced by the i th electron due to the presence of the other electrons. $v^{HF}(i)$ has two two-electron terms: the coulomb term and the exchange term. In an exact theory, the coulomb interaction is expressed as two-electron operator \mathbf{r}_{ij}^{-1} . In Hartree Fock approximation, the coulomb interaction is the total averaged potential acting on the electron in χ_a by averaging \mathbf{r}_{ij}^{-1} over all space and spin coordinates of electron j weighted by the probability $d\mathbf{x}_j |\chi_b(j)|^2$, then summing over all except a orbital. The exchange term does not have a simple classical interpretation because it arises from the antisymmetric nature of the single determinant. $v^{HF}(i)$ is mathematically expressed as the following with the first term being the coulomb term and second term being the exchange term.

$$v^{HF}(i) = \sum_{b \neq a} \left[\int d\mathbf{x}_j |\chi_b(j)|^2 r_{ij}^{-1} \right] \chi_a(i) - \sum_{b \neq a} \left[\int d\mathbf{x}_j \chi_b^*(j) \chi_a(j) r_{ij}^{-1} \right] \chi_b(i)$$

Only the ground state of closed-shell configurations with even number paired electrons will be discussed because this dissertation only deals with closed-shell molecules. The solution process is more complicated for open-shell cases.^[22] Restricted spin orbitals, which are constrained to have the same spatial function for spin up and spin down functions, are used.

Then the Hamiltonian can be expressed as

$$\mathbf{H} = \sum_{i=1}^N h(i)$$

The total electronic energy is the expectation value of $E_0 = \langle \Psi_0 | \mathbf{H} | \Psi_0 \rangle$ and given by

$$E_0 = 2 \sum_a^{N/2} h_{aa} + \sum_a^{N/2} \sum_b^{N/2} 2J_{ab} - K_{ab}$$

where h_{aa} is one-electron term, J_{ab} is coulomb integral and K_{ab} is an exchange integral.

The electronic Hamiltonian is described by the antisymmetric wavefunction formed from a set of spin orbitals with mathematical form:

$$|\Psi_0\rangle = |\chi_1 \chi_2 \cdots \chi_N\rangle$$

The procedure for solving the Hartree-Fock equation is called the self-consistent-field (SCF) method. The basic idea of the SCF method is making an initial guess at the spin orbitals, the average field seen by each electron can be calculated then the eigenvalue equation for a new set of spin orbitals can be solved. New fields can be obtained by using these new spin orbitals. The procedure is repeated until the fields no longer change. The spin orbitals are used to construct the Fock operator. The resulted total energy is not the sum of the orbital energies, which count the electron-electron interactions twice.

In 1951, Roothaan^[21] proposed representing the Hartree-Fock orbitals as linear combinations of a set of known functions, called the basis functions. The Roothaan expansion procedure allows one to find the Hartree-Fock wave function using matrix algebra, which is easy to implement on a computer and is usually used to find Hartree-Fock wave functions.

The spin orbitals can be converted to spatial orbitals ψ_i by eliminating spins. Then a set of K known basis functions can be introduced

$$\phi_\mu(\mathbf{r}) \mid \mu = 1, 2, \dots, K$$

and be used to expand the unknown molecular orbitals ψ_i in the linear expansion.

$$\psi_i = \sum_{\mu=1}^K C_{\mu i} \phi_\mu \quad i = 1, 2, \dots, K$$

It is clear that the problem of calculating the Hartree-Fock molecular orbitals reduces to the problem of calculating the set of expansion coefficients $C_{\mu i}$.

The integrated Hartree-Fock equation can be written as the following, called *Roothaan equations*.

$$\sum_v F_{\mu v} C_{vi} = \varepsilon_i \sum_v S_{\mu v} C_{vi} \quad i = 1, 2, \dots, K$$

where F is called *Fock matrix*, which is $K \times K$ Hermitian matrix, and has elements

$$F_{\mu v} = \int d\mathbf{r}_1 \phi_\mu^*(1) f(1) \phi_v(1)$$

S is called *overlap matrix*, which is also a $K \times K$ Hermitian matrix, and has elements

$$S_{\mu v} = \int d\mathbf{r}_1 \phi_\mu^*(1) \phi_v(1)$$

The *Roothaan equations* can be written more compactly as the single matrix equation.

$$\mathbf{FC} = \mathbf{SC}\varepsilon$$

where \mathbf{C} is a $K \times K$ square matrix of the expansion coefficients

$$\mathbf{C} = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1K} \\ C_{21} & C_{22} & \cdots & C_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ C_{K1} & C_{K2} & \cdots & C_{KK} \end{pmatrix}$$

ε is a diagonal matrix of the orbital energies ε_i .

We need to introduce density matrix to explicitly express the Fock matrix. For closed-shell molecule, charge density can be expressed as:

$$\begin{aligned}
\rho(\mathbf{r}) &= 2 \sum_a^{N/2} |\psi_a(\mathbf{r})|^2 = 2 \sum_a^{N/2} \psi_a^*(\mathbf{r}) \psi_a(\mathbf{r}) \\
&= 2 \sum_a^{N/2} \sum_v C_{va}^* \phi_v^*(\mathbf{r}) \sum_\mu C_{\mu a} \phi_\mu(\mathbf{r}) \\
&= \sum_{\mu\nu} \left[2 \sum_a^{N/2} C_{\mu a} C_{\nu a}^* \right] \phi_\mu(\mathbf{r}) \phi_\nu^*(\mathbf{r}) \\
&= \sum_{\mu\nu} P_{\mu\nu} \phi_\mu(\mathbf{r}) \phi_\nu^*(\mathbf{r})
\end{aligned}$$

$\rho(\mathbf{r})d\mathbf{r}$ is the probability of finding an electron in $d\mathbf{r}$ at \mathbf{r} . From the above derivation, we can define a density matrix

$$P_{\mu\nu} = 2 \sum_a^{N/2} C_{\mu a} C_{\nu a}^*$$

Given a set of basis functions, the matrix \mathbf{P} specifies the charge density.

By using density matrix, the integrated Fock matrix can be expressed as

$$\begin{aligned}
F_{\mu\nu} &= \int d\mathbf{r}_1 \phi_\mu^*(1) h(1) \phi_\nu(1) + \sum_a^{N/2} \sum_{\lambda\sigma} C_{\lambda a} C_{\sigma a}^* \left[2(\mu\nu|\sigma\lambda) - (\mu\lambda|\sigma\nu) \right] \\
&= H_{\mu\nu}^{core} + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[(\mu\nu|\sigma\lambda) - \frac{1}{2}(\mu\lambda|\sigma\nu) \right]
\end{aligned}$$

$H_{\mu\nu}^{core}$ is called core-Hamiltonian, and $(\mu\nu|\sigma\lambda)$ and $(\mu\lambda|\sigma\nu)$ are two-electron integrals.

The basis functions are normalized but not orthogonal to each other. We need to find a transformation matrix \mathbf{X} to orthogonalize the basis function. There are two ways to orthogonalize the basis set. Symmetric orthogonalization uses the inverse square root of \mathbf{S} as \mathbf{X} .

$$\mathbf{X} \equiv \mathbf{S}^{-1/2}$$

Canonical orthogonalization uses the transformation matrix

$$\mathbf{X} = \mathbf{U}\mathbf{S}^{-1/2}$$

The elements of the transformation matrix are the unitary matrix divided by the square root of the corresponding eigenvalue.

$$X_{ij} = U_{ij} / s_j^{1/2}$$

Roothaan equations are nonlinear and they need to be solved in an iterative fashion. Restricted closed-shell Hartree-Fock wave functions are solved by the SCF procedure^[23].

1. Specify molecular coordinates, atomic numbers and number of electrons, and a basis set. Calculate all the molecular integrals *overlap matrix*, *core-Hamiltonian matrix* and two-electron integrals.
2. Obtain a transformation matrix and a guess at the density matrix.
3. Calculate Fock matrix, and the transformed Fock matrix. Then diagonalize the transformed Fock matrix to obtain expansion coefficient C' and orbital energies.
4. Calculate $C = XC'$ and form a new density matrix. Determine whether the new density matrix is the same as the previous density matrix within a specified criterion. If not the same, the calculation has not converged and the calculation returns to step 3.
5. If the calculation converged, we can use the resultant expansion coefficient, density matrix and Fock matrix to calculate the energy and other expectation values of energy.

HF methods are not very accurate for NMR property calculation, however, they are fast and provide approximate values.

Møller-Plesset second order perturbation theory

The electron-electron interaction is described by an average interaction in HF.

Perturbation theory is used to estimate the electron correlation interaction. The Hamiltonian operator in perturbation theory has two parts, a reference and a perturbation. Møller-Plesset perturbation theory (MBPT),^[24] which includes a term for electron correlation energy, is widely used for, and was first applied to N -electron system by C. Møller and M.S. Plesset. It uses the Hartree-Fock Hamiltonian as a reference developed from Rayleigh-Schrödinger perturbation theory, in which the solutions to the unperturbed Schrödinger equation generate a basis set of functions. The unknown wave function can be expanded in these functions.

Møller-Plesset second order perturbation theory (MP2) is the most popular in the MBPT family because it is computationally cheap and gives the majority of the electron correlation effect. MP2 describes the correlation between pairs of electrons. Møller-Plesset fifth order perturbation theory (MP5) and higher order perturbation methods are not feasible for most molecular systems since it requires too much memory and time. MP2 usually overestimates the electron correlation effects, but it gives a better answer than Møller-Plesset third order perturbation theory MP3.

Electron correlation methods uses the HF wave function as a starting point for improvement because the HF solution usually gives about 99% of the correct answer.

The wave function can be written as

$$\Psi_i = \lambda^0 \Psi_i^{(0)} + \sum_i \lambda^i \Psi_i^{(i)}$$

where λ^0 is usually close to 1 and $\Psi_i^{(0)}$ is HF wave function. λ is a parameter determining the strength of the perturbation.

The Hamiltonian including perturbation term can be expressed as:

$$\mathbf{H} = \mathbf{H}_0 + \lambda \mathbf{H}'$$

where \mathbf{H}_0 is the Hartree-Fock Hamiltonian and $\lambda \mathbf{H}'$ is the perturbation, expressed as

$$\mathbf{H}_0 = \sum_i f(i) = \sum_i [h(i) + v^{HF}(i)]$$

The eigenvalue of following equation can be solved

$$\mathbf{H}\Psi = E\Psi$$

The eigenvalues can be expanded in Taylor series in λ ,

$$E_i = E_i^{(0)} + \sum_i \lambda^i E_i^{(i)}$$

where $E^{(i)}$ is the i th-order energy. $E^{(0)}$ is Hartree-Fock energy. The eigenvalue equation then becomes

$$(\mathbf{H}_0 + \lambda \mathbf{H}') \left(\Psi_i^{(0)} + \sum_i \lambda^i \Psi_i^{(i)} \right) = \left(E_i^{(0)} + \sum_i \lambda^i E_i^{(i)} \right) \left(\Psi_0 + \sum_i \lambda^i \Psi_i^{(i)} \right)$$

and equating the coefficients λ^i in the equation gives a i th-order perturbation equation shown in the following equations.

$$\mathbf{H}_0 \Psi^{(0)} = E^{(0)} \Psi^{(0)}$$

$$\mathbf{H}_0 \Psi^{(1)} + \mathbf{H}' \Psi^{(0)} = E^{(0)} \Psi^{(1)} + E^{(1)} \Psi^{(0)}$$

$$\mathbf{H}_0 \Psi^{(2)} + \mathbf{H}' \Psi^{(1)} = E^{(0)} \Psi^{(2)} + E^{(1)} \Psi^{(1)} + E^{(2)} \Psi^{(0)}$$

$$\mathbf{H}_0 \Psi^{(3)} + \mathbf{H}' \Psi^{(2)} = E^{(0)} \Psi^{(3)} + E^{(1)} \Psi^{(2)} + E^{(2)} \Psi^{(1)} + E^{(3)} \Psi^{(0)}$$

The above equations are solved by multiplying by Ψ_0 and using the following orthogonality relation

$$\langle \Psi_0 | \Psi_i \rangle = 0$$

The i th-order energy will be

$$E^{(0)} = \langle \Psi^{(0)} | \mathbf{H}_0 | \Psi^{(0)} \rangle$$

$$E^{(1)} = \langle \Psi^{(0)} | \mathbf{H}' | \Psi^{(0)} \rangle$$

$$E^{(2)} = \langle \Psi^{(0)} | \mathbf{H}' | \Psi^{(1)} \rangle$$

$$E^{(3)} = \langle \Psi^{(0)} | \mathbf{H}' | \Psi^{(2)} \rangle$$

MP2 theory with London orbitals remains a standard method for shielding calculations, providing accurate results at a reasonable cost.

Coupled cluster approximation

The coupled cluster (CC) approximation^[25] is one the more accurate methods in quantum chemistry for calculating dynamical electron correlation effects. The CCSDT (coupled cluster with a full treatment singles, doubles and triples) model gives excellent result but the cost of CCSDT is sufficiently high to severely restrict its application to chemical problems, and CCSD(T) (coupled cluster with a full treatment singles and doubles, and an estimate to the connected triples contributions) is the most feasible one in this family.

In the CC method the exact wave function is expressed as:

$$\Psi_{CC} = e^{\mathbf{T}} \Psi_0$$

where Ψ_0 is the ground state HF wave function and the operator $e^{\mathbf{T}}$ can be expanded by using Taylor series.

$$e^{\mathbf{T}} = 1 + \mathbf{T} + \frac{\mathbf{T}^2}{2!} + \frac{\mathbf{T}^3}{3!} + \cdots$$

\mathbf{T} is the cluster operator and defined as

$$\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 + \cdots + \mathbf{T}_n$$

where n is the number of electrons in the molecule.

In CCSD, \mathbf{T} is truncated to second order and expressed as

$$\mathbf{T} \approx \mathbf{T}_1 + \mathbf{T}_2$$

In CCSDT, \mathbf{T} is truncated to third order and expressed as

$$\mathbf{T} \approx \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3$$

The one electron excitation operator \mathbf{T}_1 is defined by

$$\mathbf{T}_1 \Psi_0 = \sum_{a=n+1}^{\infty} \sum_{i=1}^n t_i^a \Psi_i^a$$

Ψ_i^a is a singly excited Slater determinant with the occupied orbital χ_i replaced by the virtual excited orbital χ_a . t_i^a is the corresponding coefficient. The operator \mathbf{T}_1 is used to convert Ψ_0 into a linear combination of all singly excited wave functions Ψ_i^a .

The two-electron excitation operator \mathbf{T}_2 is defined by

$$\mathbf{T}_2 \Psi_0 = \sum_{b=k+1}^{\infty} \sum_{a=n+1}^{\infty} \sum_{j=i+1}^n \sum_{i=1}^{n-1} t_{ij}^{ab} \Psi_{ij}^{ab}$$

Ψ_{ij}^{ab} is a Slater determinant with the occupied orbital χ_i and χ_j replaced by the virtual orbitals χ_a and χ_b . Similarly, \mathbf{T}_1 is defined as

$$\mathbf{T}_1 = \sum_{\substack{i>j>k... \\ a>b>c}} t_{ijk...}^{abc...} \Psi_{ijk...}^{abc...}$$

The operator $e^{\mathbf{T}}$ can express the wave function as a linear combination of ground wave function and all possible excited wave functions. Once the coefficients t_{ijk}^{abc} are solved, the wave function and other properties can be found.

In CCSD(T), the full CCSDT cluster operator $e^{\mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3}$ is preserved in the creation of single and double excitation coefficients, but in the calculation of triple excitation coefficients only the \mathbf{T}_2 operator is used.^[26]

The most accurate NMR properties are obtained using coupled cluster theory. For highly accurate calculations of nuclear shielding constants inclusion of zero-point vibrational correction and extrapolation to the complete basis set limit is important. These will be discussed further in chapter 4.

Basis sets

A basis set is a set of functions that define the wave function. The introduction of basis set in *ab initio* methods is one aspect of introducing approximation. The wave function expanded in terms of a basis set is an approximation because only a complete basis set, which contains an infinite number of functions, can provide a completely accurate result. For high accuracy in calculation of shielding constants, a carefully selected basis set that properly describes the outer-core and inner-valence region is necessary. The correlation energy needs to be calculated very accurately for calculating shielding constants with highly accuracy. Pople's split-valence basis sets^[27] and Dunning's aug-cc-pVXZ and

aug-cc-pCVXZ basis sets^[28] are two kinds of popular molecular basis sets that use Gaussian functions.

Basis sets were first developed by J.C.Slater. A minimal basis set for a molecular SCF calculation, consists of a single basis function for each inner-shell atomic orbital and each valence-shell atomic orbital of each atom, describes only the most basic aspects of the orbitals. Slater Type Orbitals (STO)^[29], the simplest basis sets, and Gaussian Type Orbitals (GTO)^[30], modified STO, are two kinds of minimal basis sets. STO equation can be expressed as following:

$$STO = \frac{\xi^3}{\pi^{0.5}} e^{-\xi r}$$

where ξ is orbital exponent and r is radius in angstroms.

The GTO squares the “ r ”. Equation will be more accurate with more Gaussian equations.

An extended basis set is a set provides much more detailed description of the orbitals. Split-valence basis sets and Dunning’s basis sets are extended basis sets.

Split-valence basis sets introduce an extra basis function to describe each valence orbital. The 4-31 G basis set is recognized to give reasonable agreement for theoretical equilibrium geometries and experiment. The 4-31 G basis set uses a sum of four Gaussian functions to describe the inner shell; and each valence shell is split into inner and outer parts, which is described by three and one Gaussian functions^[27d]. By using more Gaussian functions for the inner shell, the calculated energy will approach the Hartree-Fock limit^[27a]. 6-31G* and 6-31G** can improve the calculated molecular properties by

adding functions to allow for a displacement of the center of electron density from the individual nuclei. This is called a polarization function^[27c]. Polarization functions are added to split-valence basis set by using a *d*-type function for the first row atoms and *p*-type Gaussian functions for hydrogen^[27e]. Diffuse functions in *sp* space is important in the description of anions, hydrogen bonding and lone pair electrons^[27f].

Dunning's basis sets uses the contraction of Gaussian functions, the exponents of the functions optimized in atomic correlated calculation^[28a]. These basis sets provide an excellent description of molecular correlation effects.

The *sp* correlation sets are obtained by augmenting the atomic orbitals with the most diffuse primitive functions in the set. Correlation consistent sets are proposed for all the first row atoms as^[28c]:

	Primitive	Contracted	Polarization set
cc-pVDZ	9s4p	3s2p	1d
cc-pVTZ	10s5p	4s2p	2d1f
cc-pVQZ	12s6p	5s4p	3d 2f 1g

A more diffuse character of correlation-consistent basis sets can be obtained by adding augmented functions. A set of primitive *s* and *p* functions was added to the *sp* sets of each of the correlation-consistent basis sets. aug-cc-pVDZ have 23 basis functions with *10s5p2d* functions contracted to *4s3p2d*. aug-cc-pVTZ have 46 basis functions with

11s6p3d2f functions contracted to 5s4p3d2f. aug-cc-pVQZ have 80 basis functions with 13s7p4d3f2g functions contracted to 6s5p4d3f2g^[28d].

It is well known that the properties of atoms and molecules are largely determined by the valence electrons. For very highest accuracy in the calculation of spectroscopic constants^[31], the effect of core-valence correlation needs to be considered. Core and core-valence correlation effects account for a small but significant source of error in a quantum chemical description of atomic and molecular systems. This source of error has been addressed by adding 1s1p for cc-pVDZ, 2s2p1d for cc-pVTZ and 3s3p2d1f for cc-pVQZ for high accuracy calculation^[28e].

The extrapolation of the calculated shielding to the complete basis set (CBS) limit is obtained by fitting the calculated results with two or three parameter reciprocal cubic or exponential functions^[32].

$$\sigma(x) = \sigma(\infty) + A / x^3$$

$$\sigma(x) = \sigma(\infty) + A * \exp(-x / B)$$

Both NMR and quantum chemistry basics are briefly reviewed here. They will be used in the following chapters.

References

- [1] aRabi, I., *Physical Review* **1937**, 51, 0652-0654; bRabi, I., J. R. Zacharias, S. Millman, P. Kusch, *Physical Review* **1938**, 53, 318-318.
- [2] E. M. Purcell, H. C. Torrey, R. V. Pound, *Physical Review* **1946**, 69, 37-38.
- [3] aF. Bloch, *Physical Review* **1946**, 70, 460-474; bF. Bloch, W. W. Hansen, M. Packard, *Physical Review* **1946**, 69, 127-127.
- [4] W. G. Proctor, F. C. Yu, *Physical Review* **1950**, 77, 717-717.
- [5] W. C. Dickinson, *Physical Review* **1950**, 77, 736-737.
- [6] J. T. Arnold, S. S. Dharmatti, M. E. Packard, *The Journal of Chemical Physics* **1951**, 19, 507-507.
- [7] R. R. Ernst, W. A. Anderson, *Review of Scientific Instruments* **1966**, 37, 93-+.
- [8] R. R. Ernst, W. A. Anderson, *Review of Scientific Instruments* **1966**, 37, 93-102.
- [9] J. Jeener, **1971**.
- [10] M. Mehring, *Principles of High Resolution NMR in Solids*, **1983**.
- [11] N. F. Ramsey, E. M. Purcell, *Physical Review* **1952**, 85, 143-144.
- [12] T. Helgaker, M. Jaszunski, K. Ruud, *Chemical Reviews* **1999**, 99, 293-352.
- [13] W. H. Flygare, *Chemical Reviews* **1974**, 74, 653-687.
- [14] Hindermä-Dekker, C. D. Cornwell, *Journal of Chemical Physics* **1968**, 48, 4148-8.
- [15] A. Pines, M. G. Gibby, J. S. Waugh, *The Journal of Chemical Physics* **1973**, 59, 569-590.
- [16] J. Schaefer, E. O. Stejskal, *Journal of the American Chemical Society* **1976**, 98, 1031-1032.
- [17] aM. S. Gordon, M. W. Schmidt, *Advances in electronic structure theory: GAMESS a decade later*, **2005**; bM. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *Journal of Computational Chemistry* **1993**, 14, 1347-1363.
- [18] M. J. Frisch, G. W. Trucks, H. B. Schlegel, et al., *GAUSSIAN 09, Revision A.02* **2009**.
- [19] J. G. J. F. Stanton, M. E. Harding, P. G. Szalay, see <http://www.cfour.de>.
- [20] aD. R. Hartree, *Proceedings of the Cambridge Philosophical Society* **1928**, 24, 89-110; bD. R. Hartree, *Proceedings of the Cambridge Philosophical Society* **1928**, 24, 111-132; cD. R. Hartree, *Proceedings of the Cambridge Philosophical Society* **1928**, 24, 426-437.
- [21] C. C. J. Roothaan, *Reviews of Modern Physics* **1951**, 23, 69-89.
- [22] I. N. LEVVINE, *Quantum chemistry*, **2000**.
- [23] A. O. Szabo, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronoc Structure Theory*, Dover, **1996**.
- [24] C. Møller, M. S. Plesset, *Physical Review* **1934**, 46, 618-622.
- [25] J. Čížek, *The Journal of Chemical Physics* **1966**, 45, 4256-4266.
- [26] M. Urban, J. Noga, S. J. Cole, R. J. Bartlett, *The Journal of Chemical Physics* **1985**, 83, 4041-4046.
- [27] aW. J. Hehre, Ditchfield, R. J. A. Pople, *Journal of Chemical Physics* **1972**, 56, 2257-8; bJ. S. Binkley, J. A. Pople, W. J. Hehre, *Journal of the American Chemical Society* **1980**, 102, 939-947; cM. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. Defrees, J. A. Pople, *Journal of Chemical Physics*

- 1982**, 77, 3654-3665; dR. Ditchfield, W. J. Hehre, J. A. Pople, *The Journal of Chemical Physics* **1971**, 54, 724-728; eP. C. Hariharan, J. A. Pople, *Theoret. Chim. Acta* **1973**, 28, 213-222; fM. J. Frisch, J. A. Pople, J. S. Binkley, *The Journal of Chemical Physics* **1984**, 80, 3265-3269.
- [28] aT. H. Dunning, *Journal of Chemical Physics* **1970**, 53, 2823-&; bT. H. Dunning, *Journal of Chemical Physics* **1971**, 55, 716-&; cT. H. Dunning, *Journal of Chemical Physics* **1989**, 90, 1007-1023; dR. A. Kendall, T. H. Dunning, R. J. Harrison, *Journal of Chemical Physics* **1992**, 96, 6796-6806; eD. E. Woon, T. H. Dunning, *Journal of Chemical Physics* **1995**, 103, 4572-4585.
- [29] J. C. Slater, *Physical Review* **1930**, 36, 0057-0064.
- [30] S. F. Boys, *Proceedings of the Royal Society of London Series a-Mathematical and Physical Sciences* **1950**, 200, 542-554.
- [31] H. Partridge, C. W. Bauschlicher, S. P. Walch, B. Liu, *The Journal of Chemical Physics* **1983**, 79, 1866-1873.
- [32] F. Jensen, *Introduction to Computational Chemistry*, Second Edition ed., Wiley, **2007**.

Chapter 1

Continuous and Smooth Potential Energy Surface for Cavitation, Dispersion and Repulsion Energies in Polarizable Continuum Model

This work has been published

Y. Wang, H. Li, *The Journal of Chemical Physics* **2009**, *131*, 206101

Abstract

Rigorously continuous and smooth potential energy surfaces, as well as exact analytic gradients, are obtained for cavitation, dispersion and repulsion energies for the continuum solvation model using the tessellation scheme, fixed points with variable areas (FIXPVA).

1.1 INTRODUCTION

Cavitation energy and dispersion-repulsion energies, which are nonelectrostatic terms included in the Gibbs energy of solvation, cannot be neglected for accurately evaluating solvent effects^[1]. Pierrotti and Claverie developed a method for evaluating cavitation energy based on Reiss's scaled particle theory^[2]. Claverie derived statistically averaged dispersion-repulsion energy expression based on atom-atom potentials^[3]. These two methods were implemented in GAMESS using the GEPOLE tessellation scheme^[4]. Numerical gradients of the cavitation, dispersion and repulsion energies were also implemented. However, due to the intrinsic discontinuity of the tessera coordinates as functions of molecular geometry in GEPOLE, the potential energy surfaces are not continuous and smooth, and geometry optimization processes often converge with difficulty.

Recently Li's group developed a new tessellation scheme called Fixed Point with Variable Area (FIXPVA)^[5], and obtained rigorously continuous and smooth potential energy surfaces, as well as exact analytic gradients for the electrostatic solvation energy in homogeneous and heterogeneous COSMO and CPCM.

In this work, the FIXPVA tessellation scheme was extended to treat cavitation, dispersion and repulsion energies, and rigorously continuous and smooth potential energy surfaces, as well as exact gradients, were obtained.

1.2 THEORY

Cavitation, dispersion and repulsion energy

Based on the scaled particle theory originally developed by Reiss, Frisch, Helfand and Lebowitz,^[6] Pierotti formulated a practical method and derived an expression for the molar Gibbs free energy of cavitation:^[7]

$$G_{cav} = \sum_i \frac{a_i}{4\pi R_i^2} [K_0 + K_1(R_i + R_s) + K_2(R_i + R_s)^2] \quad (1)$$

where i runs over all the tesserae, a_i is the area of tessera i , R_i is the radius of the sphere associated with tessera i , and R_s is the assumed radius of the solvent molecule. The meaning and derivation of the parameters K_0 , K_1 , K_2 can be found in the literature^[8].

The statistically averaged dispersion and repulsion energies based on a Buckingham potential is^[3b, 9]

$$\rho \sum_N \sum_M \sum_i \left\{ \left[\left(\frac{1}{\alpha_{MN} |\mathbf{r}_M - \mathbf{r}_i|} + \frac{2}{\alpha_{MN}^2 |\mathbf{r}_M - \mathbf{r}_i|^2} + \frac{2}{\alpha_{MN}^3 |\mathbf{r}_M - \mathbf{r}_i|^3} \right) \beta_{MN} \exp(-\alpha_{MN} |\mathbf{r}_M - \mathbf{r}_i|) - \frac{C_{MN}}{|\mathbf{r}_M - \mathbf{r}_i|^6} \right] \cdot \frac{R_i^2 + |\mathbf{r}_M - \mathbf{r}_i|^2 - |\mathbf{r}_M - \mathbf{r}_I|^2}{2R_i} \cdot a_i \right\} \quad (2)$$

where N runs over all solvent atoms, M runs over all solute atoms, and i runs over all tesserae. ρ is the numerical density of the solvent. a_{MN} and b_{MN} are the two parameters in the exponential repulsion potential between solute atom M and solvent atom N ; C_{MN} is the dispersion coefficient for solute atom M and solvent atom N . \mathbf{r}_M and \mathbf{r}_i are the coordinates of solute atom M and tessera i . R_i and \mathbf{r}_I are, respectively, the radius and center coordinate of the sphere associated with tessera i . a_i is the area of tessera i .

Similarly, the statistically averaged dispersion-repulsion energy based on the Lennard-Jones potential is:

$$\sum_M \sum_i \left(\frac{B_M}{|\mathbf{r}_M - \mathbf{r}_i|^{12}} - \frac{C_M}{|\mathbf{r}_M - \mathbf{r}_i|^6} \right) \frac{R_i^2 + |\mathbf{r}_M - \mathbf{r}_i|^2 - |\mathbf{r}_M - \mathbf{r}_I|^2}{2R_i} a_i \quad (3)$$

where M runs over all solute atoms, i runs over all tesserae. and B_M and C_M are the repulsion and dispersion parameters for solute atom M and the solvent molecule. All the other variables have the same meaning as those in Eq. (2).

The derivative of cavitation and dispersion-repulsion energies with respect to a nuclear coordinate x is trivial. In general, it involves the derivatives of the tessera coordinates \mathbf{r}_i^x and areas a_i^x . The derivatives of the solute atomic coordinates are zero if they are not x . Other parameters in Eqs (1), (2) and (3) are constants.

FIXPVA tessellation

The analytic derivative of the tessera areas and coordinates can be obtained with the FIXPVA tessellation procedure. FIXPVA has the following features: (1) each atom is assigned with a sphere and no additional spheres are used; (2) the surface of each sphere is divided into 60, 240 or 960 initial tesserae; (3) the center of each tessera is fixed relative to the associated atom, but its area is scaled by switching functions of its distances to neighboring spheres.

Appropriate values for parameters m_1 , m_2 , n_1 and n_2 in switching functions can ensure continuous analytical gradients. Su and Li assigned 0.02 Å, 0.3 Å, 1.0 Å and 1.5 Å separately for m_1 , m_2 , n_1 and n_2 .

Compared to GEPOL, FIXPVA produces ~10% less surface area. In order to get a similar energy for the FIXPVA method, it is necessary to adjust the switching functions so the results from GEPOL are reproduced.

Based on extensive tests, m_1 , m_2 , n_1 and n_2 were 0.02, 0.30, 0.50, 1.00 and R_B was reduced by 0.16 Å for the Pierotti-Claverie cavitation energy. m_1 , m_2 , n_1 , n_2 were picked as 0.02, 0.30, 1.00, 1.50 and R_B was reduced by 0.11 Å for dispersion and repulsion.

1.3 COMPUTATIONAL METHODOLOGY

On the basis of previous implementations, the cavitation, dispersion and repulsion energies and nuclear gradients were implemented in GAMESS using the FIXPVA tessellation scheme. New values of parameters m_1 , m_2 , n_1 , n_2 and R_B were implemented specifically for cavitation, repulsion and dispersion also. Water with ϵ 78.39 was chosen as solvent.

In cavitation, dispersion and repulsion calculations, the default parameters for cavitation, dispersion and repulsion were used. Sphere with radii of 0, 1.77, 1.68, 1.59 and 2.10 Å were used for H, C, N, O and S atoms, respectively, to define the molecular cavity and surface tesserae in the FIXPVA tessellation scheme. For CPCM electrostatic solvation energy calculations, these radii were scaled by 1.20. For cavitation energy calculation, they are used as defined. For dispersion-repulsion energy calculation, the radii of the atoms of the solvent molecule (in this work, H, and O of water) were added to the solute atomic radii to define the molecular surface tesserae.

1.4 RESULTS AND DISCUSSION

Energies

Table 1.1 presents cavitation, dispersion and repulsion energies calculated with GEPOL scheme (60 initial tesserae per sphere) and FIXPVA scheme (240 initial tesserae per sphere) for 11 molecules. The FIXPVA results are close to the GEPOL results. For cavitation energies, the maximum deviation is 0.21 kcal/mol and root-mean-square deviation is 0.08 kcal/mol. Most molecules show deviations that are smaller than 0.1 kcal/mol. The maximum deviations of the FIXPVA repulsion and dispersion energies are 0.19 and 0.75 kcal/mol, respectively, from the GEPOL values, with RMSD being 0.09 and 0.37 kcal/mol, respectively.

Gradients

Table 1.2 presents the analytic and numerical gradients of the total FIXPVA-CPCM/RHF/631G* energy, including cavitation, dispersion and repulsion terms, for acetate at the geometry optimized with the gas phase RHF/6-31G*. The numerical gradients were obtained by double displacements (backward and forward) with a step size of 0.001 au for each step. The maximum deviation between the numerical and analytic gradients is 5×10^{-7} au, with a root-mean-square deviation of 2×10^{-7} .

Rotational variance

A molecule may have different energies at different orientations because the GEPOL and FIXPVA tessellations are not rotationally invariant. Twenty different orientations of acetate were used to test the rotational variances of the cavitation, dispersion and repulsion energies. Using GEPOL, the maximum rotational variances observed for cavitation, repulsion and dispersion energies are all less than 0.1 kcal/mol

when 60 initial tesserae per sphere are used. Using FIXPVA, the maximum variances observed for cavitation, repulsion and dispersion are 0.83, 0.25 and 1.12 kcal/mol, respectively, if 60 initial tesserae per sphere were used, and are 0.21, 0.08 and 0.42 kcal/mole, respectively, if 240 initial tesserae per sphere were used. Therefore, 240 initial tesserae per sphere are recommended for general use.

Geometry optimization

The 2,5-diketopiperazine-3-acetate ($C_6H_7O_4N_2$) anion was optimized with the CPCM/RHF/6-31G* method, in which the Pierotti-Claverie cavitation energy and the Claverie dispersion-repulsion energy were included. The FIXPVA scheme was used for the CPCM electrostatic solvation energy calculation, but either GEOPOL or FIXPVA was used for the cavitation, dispersion and repulsion energy calculation. The energy and root-mean-square gradient (RMSG) profiles are shown in Figure 1.1. Using GEOPOL with 60 initial tesserae per sphere, the energy in the geometry optimization process fluctuates, reflecting the fact that the potential energy surface is bumpy; the RMSG shows large and random fluctuations, with no sign of convergence. In fact the geometry optimization did not converge in 250 steps (only 100 steps are shown in Figure 2) to the criteria that the maximum gradient be smaller than 5×10^{-5} au and the RMSG be smaller than 1.66×10^{-5} au. Using FIXPVA with 240 initial tesserae per sphere, the maximum gradient and RMSG smoothly decreases and falls below the convergence criterion at the 36th step.

REFERENCES

- [1] (a) J. Tomasi, B. Mennucci, R. Cammi, *Chemical Reviews* **2005**, *105*, 2999-3093; (b) H. S. Paulo F. B. Goncalves, *Journal of Computational Chemistry* **2003**, *24*, 1758-1765; (c) H. H. Uhlig, *The Journal of Physical Chemistry* **1937**, *41*, 1215-1226.
- [2] (a) J. Tomasi, M. Persico, *Chemical Reviews* **1994**, *94*, 2027-2094; (b) R. A. Pierotti, *Chemical Reviews* **1976**, *76*, 717-726; (c) H. Reiss, D. M. Tully-Smith, *The Journal of Chemical Physics* **1971**, *55*, 1674-1689.
- [3] (a) J. T. F. Floris, *Journal of Computational Chemistry* **1989**, *10*, 616-627; (b) J. T. F. M. Floris, J. L. Pascual-Ahuir, *Journal of Computational Chemistry* **1991**, *12*, 784-791.
- [4] (a) J. L. Pascual-Ahuir, E. Silla, I. Tunon, *Journal of Computational Chemistry* **1994**, *15*, 1127-1138; (b) E. Silla, I. Tunon, J. L. Pascual-Ahuir, *Journal of Computational Chemistry* **1991**, *12*, 1077-1088; (c) E. Silla, F. Villar, O. Nilsson, J. L. Pascual-Ahuir, O. Tapia, *Journal of Molecular Graphics* **1990**, *8*, 168-172; (d) J. L. Pascual-Ahuir, E. Silla, *Journal of Computational Chemistry* **1990**, *11*, 1047-1060.
- [5] P. Su, H. Li, *The Journal of Chemical Physics* **2009**, *130*, 074109.
- [6] H. Reiss, H. L. Frisch, E. Helfand, J. L. Lebowitz, *The Journal of Chemical Physics* **1960**, *32*, 119-124.
- [7] R. A. Pierotti, *Chemical Review* **1976**, *76*, 717-726.
- [8] J. Langlet, P. Claverie, J. Caillet, A. Pullman, *The Journal of Physical Chemistry* **1988**, *92*, 1617-1631.
- [9] R. Cammi, J. Tomasi, *Journal of Chemical Physics* **1994**, *101*, 3888-3897.

Table 1.1 Cavitation, dispersion and repulsion energies (kcal/mol) calculated for 11 molecules.

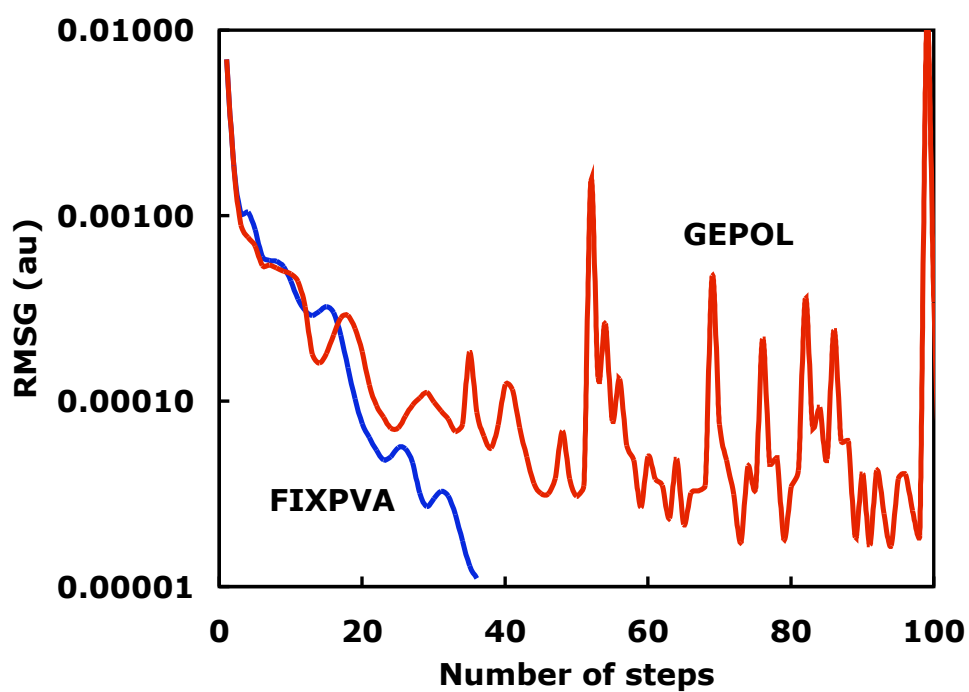
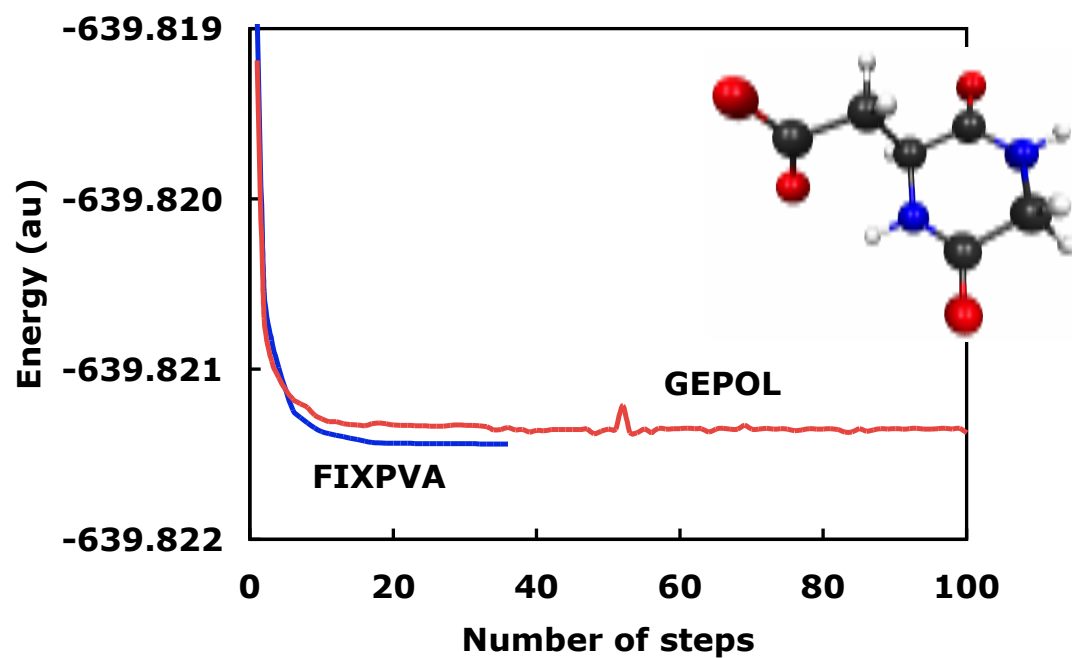
Energies	Cavitation		Repulsion		Dispersion	
	GEPOL	FIXPVA	GEPOL	FIXPVA	GEPOL	FIXPVA
C ₄ H ₅ O ₂ N ₂ CH ₂ COO ⁻ (a)	20.56	20.51	4.61	4.63	-22.51	-22.32
C ₆ H ₅ COO ⁻	16.19	16.25	3.89	4.04	-18.49	-19.24
CH ₃ NHCOCH ₂ COO ⁻	15.76	15.74	3.91	3.86	-17.62	-17.19
Pyrrolidine-2-formate, C ₄ H ₈ NHCOO ⁻	15.24	15.16	4.83	4.95	-19.49	-19.62
NH ₂ COCH ₂ COO ⁻	13.92	13.91	2.97	3.00	-14.93	-14.94
C ₆ H ₆	11.87	12.08	4.26	4.45	-16.34	-16.9
1-H-imidazole, C ₃ H ₄ N ₂	10.05	10.15	3.22	3.31	-12.99	-13.54
CH ₃ COO ⁻	9.46	9.47	2.63	2.58	-10.82	-10.65
CH ₃ S ⁻	8.32	8.27	2.12	2.11	-8.10	-7.94
CH ₃ O ⁻	6.21	6.25	2.68	2.71	-8.35	-8.42
CH ₃ NH ₂	6.60	6.55	3.38	3.46	-10.31	-10.43
Maximum deviation		0.21		0.19		0.75
RMSD		0.08		0.09		0.37

(a) 2,5-diketopiperazine-3-acetate

TABLE 1.2 Gradients (au) of the total CPCM/RHF/6-31G* energy for CH₃COO⁻.

Coordinates	Numerical	Analytic
C1X	-0.0044595	-0.0044593
C1Y	0.0011603	0.0011600
C1Z	-0.0051335	-0.0051334
C2X	0.0065204	0.0065207
C2Y	-0.0009628	-0.0009625
C2Z	0.0251118	0.0251115
O3X	-0.0005102	-0.0005107
O3Y	0.0001747	0.0001748
O3Z	-0.0105796	-0.0105795
O4X	-0.0008810	-0.0008808
O4Y	-0.0003624	-0.0003624
O4Z	-0.0079562	-0.0079562
H5X	0.0012934	0.0012933
H5Y	0.0000114	0.0000114
H5Z	-0.0007937	-0.0007937
H6X	-0.0010108	-0.0010107
H6Y	0.0005702	0.0005701
H6Z	-0.0003092	-0.0003091
H7X	-0.0009523	-0.0009522
H7Y	-0.0005915	-0.0005914
H7Z	-0.0003396	-0.0003396
Max		0.0000005
deviation		
RMSD		0.0000002

Figure 1.1 The energy and root mean square gradient (RMSG) profiles in CPCM/RHF/6-31G* geometry optimization of 2,5-diketopiperazine-3-acetate ($C_6H_7O_4N_2$),



Chapter 2

Energy Gradients in Combined Time-Dependent Density Functional Theory and Conductorlike Polarizable Continuum Model Method

This work has been published as

Y. Wang, L. Hui, *The Journal of Chemical Physics* **2010**, *133*, 034108

Abstract

The analytical gradient of the combined conductorlike polarizable continuum model (CPCM) and time-dependent density functional theory (TDDFT) method is derived and implemented. Nuclear gradients computed using CPCM/TDDFT are as accurate as those obtained by gas phase TDDFT methods.

2.1 Introduction

Using the Runge-Gross theorem^[1], about twenty years ago Casida *et al* developed time dependent density functional theory (TDDFT) for calculating excitation energies^[2]. Since then, TDDFT has been widely used to study excited state properties due to its low computational cost and relatively high accuracy, especially, for valence excited states^[3]. TDDFT method was also derived variationally by the Wang group^[4]. It is well known that analytical gradients can be used to efficiently perform geometry optimization and dynamics simulations. Caillie and Amos^[5], and Furche and Ahlrichs^[6] derived the analytic gradients of the gas phase TDDFT excitation energy, with the TDDFT relaxed density being determined by the **Z**-vector method originally developed by Handy and Schaefer^[7].

Solvent effects must be considered in quantum chemical calculations of molecular geometries and properties for condensed phases. Explicit molecular mechanical solvation models, which can provide specific intermolecular interactions, but are more expensive, and implicit continuum solvation models, which are more efficient, but omit specific solvent interactions, are two essential methods for solvation effect calculations. A polarizable implicit model can describe the fast electron density polarization of solvent by using induced dipoles, as well as the slow molecular polarization due to solvent molecule reorientation. The polarizable continuum model (PCM), including D-PCM^[8], integral equation formalism PCM (IEFPCM)^[9] and conductor like PCM (CPCM)^[10], are popular and efficient solvation models. In PCM, solvents are treated as bulk media characterized by dielectric constants and solute is described by a quantum mechanical

method. The electrostatic polarization interaction between the solute and the solvent is determined through the induced apparent surface charge method. The Li group obtained rigorously continuous and smooth potential energy surface and exact gradient for CPCM using the fixed points with variable areas (FIXPVA)^[11]. Continuous and smooth potential energy surface were obtained by other groups^[12] for various other versions of PCM

A few polarizable solvation models for excited states have also been developed^[13]. However, analytical gradients are not available for these methods so efficient geometry optimization and molecular dynamics simulation cannot be performed. Cossi and Barone developed the combined TDDFT and PCM method^[14], for which analytic gradients were derived and implemented by the Scalmani group^[6a, 15]. In the PCM/TDDFT method, solvent electronic polarization is described using the optical dielectric constant while the total polarization (electronic + molecular) is described using the static dielectric constant.

In this work, the analytical gradients for the TDDFT/CPCM excitation energy were derived and implemented by using FIXPVA. CPCM is considered because it is easier to derive gradients although IEFPCM^[16] is more rigorous from an electrostatic interaction point. This is the first time that rigorously continuous and smooth CPCM/TDDFT potential energy surfaces have been obtained.

This chapter is organized as following: First CPCM, TDDFT and CPCM/TDDFT theory will be reviewed; then the implementation of the CPCM/TDDFT code and the general computational method will be described; Finally the application to p-nitroaniline, acetone and PYP is discussed.

2.2 Theory

Conductorlike Polarizable Continuum Model (CPCM)

In CPCM, solvent is represented by a dielectric continuum. Induced surface charges initially determined by the solute are distributed on the solute cavity surface. By using the self-consistent formalism, the induced surface charges are determined through the interaction with electrostatic potentials. Induced point charges at the surface tesserae are used to represent the continuous distribution of surface charges by solving the Poisson equation. The induced point charges \mathbf{q} are determined by electrostatic potentials \mathbf{V} at the solute cavity tesserae according to the following linear response equation:

$$\mathbf{C} \cdot \mathbf{q} = -k\mathbf{V} \quad (1)$$

where k is a function of dielectric constant representing the ratio of solute charge screened by solvent, $k = (\epsilon - 1)/\epsilon$ in CPCM.

The elements of matrix \mathbf{C} are:

$$\mathbf{C}_{ii} = 1.07\sqrt{4\pi/a_i}, \quad \mathbf{C}_{ij} = 1/|\mathbf{r}_i - \mathbf{r}_j| \quad (2)$$

with a_i being the area and \mathbf{r}_i being the center coordinates of tessera i .

The electrostatic potential \mathbf{V} , which is created by the solute nuclei and electrons, can be written as:

$$\mathbf{V} = \mathbf{V}^{nuc} + \mathbf{V}^{ele} \quad (3)$$

with

$$\mathbf{V}^{nuc} = \sum_{\alpha} \frac{Z_{\alpha}}{|\mathbf{r}_{\alpha} - \mathbf{r}|} \quad (3a)$$

$$\mathbf{V}^{ele} = \sum_{\rho\sigma} P_{\mu\nu} V_{\mu\nu} \quad (3b)$$

$$\mathbf{V}_{\mu\nu} = - \left\langle \mu \left| \frac{1}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle \quad (3c)$$

where $\mathbf{V}_{\mu\nu}$ is the basis set potential; Z_α and \mathbf{r}_α are the nuclear charge and coordinates of atom α ; ρ and σ , as well as μ and ν used in this paper, are Gaussian type basis functions; $P_{\mu\nu}$ is the density matrix and \mathbf{r}_e is the electronic coordinate.

The solvation operator \mathbf{V}^{sol} , which can be incorporated into the gas phase Kohn-Sham operator, is expressed in molecular orbital basis set as:

$$\mathbf{V}_{pq}^{sol} = - \left\langle p \left| \frac{\mathbf{q}}{|\mathbf{r}_e - \mathbf{r}_i|} \right| q \right\rangle \quad (4)$$

where p and q represent basis functions.

Gas phase TDDFT

The central equation in the gas phase TDDFT method is

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \omega \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \quad (6)$$

where \mathbf{X} and \mathbf{Y} are the first-order or linear response amplitudes in the excitation and de-excitation processes; ω is the excitation energy, \mathbf{A} and \mathbf{B} matrices are the orbital rotation Hessian with the following elements,

$$A_{ia,jb} = \delta_{ij} \delta_{ab} (\epsilon_a - \epsilon_i) + K_{ia,jb} \quad , \quad B_{ia,jb} = K_{ia,bj} \quad (7)$$

with i and j representing the occupied spin orbitals, a and b representing the virtual spin orbitals, ϵ representing the spin orbital energies, and \mathbf{K} representing the coupling matrix,

which contains the DFT exchange-correlation term $f_{ai,bj}^{XC}$, and Hartree-Fock exchange terms in hybrid DFT methods. Eq (6) becomes the time-dependent Hartree-Fock (TDHF) equation if the \mathbf{K} matrix contains pure Hartree-Fock exchange terms.

Eq (6) has solutions corresponding to various gas phase excited states. The TDDFT n th excitation energy ω (we omitted the n for the n th state for simplicity) is

$$\omega = \frac{1}{2} \langle \mathbf{X} + \mathbf{Y} | \mathbf{A} + \mathbf{B} | \mathbf{X} + \mathbf{Y} \rangle + \frac{1}{2} \langle \mathbf{X} - \mathbf{Y} | \mathbf{A} - \mathbf{B} | \mathbf{X} - \mathbf{Y} \rangle \quad (8)$$

The total energy of the excited state is simply the sum of the ground state DFT energy and the TDDFT excitation energy ω .

Furche and Ahlrichs^[6] have developed an efficient “relaxed density” method for evaluating the gradients of the gas phase TDDFT excitation energy ω , with the TDDFT relaxed density being determined by the \mathbf{Z} -vector method originally developed by Handy and Schaefer^[7]. According to Furche and Ahlrichs, the first derivative of the TDDFT excitation energy ω with respect to an atomic coordinate x is

$$\begin{aligned} \omega^x &= \frac{1}{2} \langle \mathbf{X} + \mathbf{Y} | (\mathbf{A} + \mathbf{B})^x | \mathbf{X} + \mathbf{Y} \rangle + \frac{1}{2} \langle \mathbf{X} - \mathbf{Y} | (\mathbf{A} - \mathbf{B})^x | \mathbf{X} - \mathbf{Y} \rangle \\ &= \sum_{\mu\nu} P_{\mu\nu}^{\Delta} h_{\mu\nu}^x - \sum_{\mu\nu} W_{\mu\nu}^{\Delta} S_{\mu\nu}^x + \sum_{\mu\nu} P_{\mu\nu}^{\Delta} (V_{\mu\nu}^{XC})^x + \sum_{\mu\nu\rho\sigma} \Gamma_{\mu\nu\rho\sigma}^{\Delta} (\mu\nu | \rho\sigma)^x \\ &\quad + \sum_{\mu\nu\rho\sigma} (X + Y)_{\mu\nu} (X + Y)_{\rho\sigma} (f_{\mu\nu\rho\sigma}^{XC})^x \end{aligned} \quad (9)$$

where μ , ν , ρ and σ denote basis functions, \mathbf{h} and \mathbf{S} are the basis set one-electron and overlap integral matrices (derivative of density matrices), and \mathbf{W}^{Δ} is the energy-weighted difference density matrix, which absorbs all density matrix involved terms. $\mathbf{\Gamma}^{\Delta}$ is the two-particle difference density matrix, and f^{XC} is the Hartree-Fock or DFT exchange-correlation kernel in the adiabatic approximation. The relaxed one-particle difference

density matrix \mathbf{P}^Δ is the sum of the unrelaxed difference density matrix \mathbf{T}^Δ and the \mathbf{Z} matrix from the \mathbf{Z} -vector method^[6a]:

$$\mathbf{P}^\Delta = \mathbf{T}^\Delta + \mathbf{Z} \quad (10)$$

The expression of \mathbf{T}^Δ , \mathbf{Z} and \mathbf{W}^Δ can be found in the literature^[6a].

TDDFT/CPCM

In general, CPCM affects the TDDFT energies in two ways. First, the CPCM induced charges affect the ground state DFT spin orbitals and their energies, thus implicitly affect the TDDFT results. Second, the CPCM induced electronic charges (only due to solute electrons) explicitly affect the \mathbf{A} and \mathbf{B} matrices in Eq (6).

$$\begin{aligned} A_{ia,jb} &= \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + K_{ia,jb} + E_{ia,jb}^{sol} \\ B_{ia,jb} &= K_{ia,bj} + E_{ia,jb}^{sol} \end{aligned} \quad (11)$$

where the CPCM polarization term $E_{ia,jb}^{sol}$ is the sum of the electrostatic interaction energy between the potential \mathbf{V}_{ia} created by the electron density $\psi_i\psi_a$ and the CPCM charge \mathbf{q}_{bj} induced by the electron density $\psi_j\psi_b$:

$$E_{ia,jb}^{sol} = \mathbf{V}_{ia} \mathbf{q}_{jb} \quad (12)$$

For TDDFT/CPCM calculation of solvent effects on nonequilibrium vertical transmission processes, the static dielectric constant of the solvent should be used in CPCM for the initial equilibrium state, but only the optical dielectric constant should be used in CPCM for the final nonequilibrium state. It is convenient and a good approximation to separate the total induced CPCM surface charge \mathbf{q} into two components, a fast response charge \mathbf{q}_{fast} and a slow response charge \mathbf{q}_{slow} , and link them

to the optical dielectric constant ϵ_{opt} of the solvent, which is simply the square of the refractive index^[14, 17]. For TDDFT/CPCM calculation of solvent effect on equilibrium properties of excited molecules, only the static dielectric constant should be used for both the initial and final states.

According to Tomasi et al^[6a, 15], the gradients of the TDDFT/CPCM excitation energy is:

$$\begin{aligned} \omega^x = & \sum_{\mu\nu} P_{\mu\nu}^{\Delta} h_{\mu\nu}^x - \sum_{\mu\nu} W_{\mu\nu}^{\Delta} S_{\mu\nu}^x + \sum_{\mu\nu} P_{\mu\nu}^{\Delta} (V_{\mu\nu}^{XC})^x + \sum_{\mu\nu\rho\sigma} \Gamma_{\mu\nu\rho\sigma}^{\Delta} (\mu\nu|\rho\sigma)^x \\ & + \sum_{\mu\nu\rho\sigma} (X+Y)_{\mu\nu} (X+Y)_{\rho\sigma} (f_{\mu\nu\rho\sigma}^{XC})^x \\ & + \sum_{\mu\nu} P_{\mu\nu}^{\Delta} (V_{\mu\nu}^{sol})^x + \sum_{\mu\nu\rho\sigma} (X+Y)_{\mu\nu} (X+Y)_{\rho\sigma} (E_{\mu\nu,\rho\sigma}^{sol})^x \end{aligned} \quad (13)$$

where the notations have the same meaning as presented before.

The first five terms in Eq (13) have exactly the same the forms as those in Eq (9) for the gas phase TDDFT excitation energy gradient, but with \mathbf{P}^{Δ} and \mathbf{W}^{Δ} perturbed by the CPCM effect.

The sixth term in Eq (13) involves the derivative of the CPCM solvation interaction energy, and can be expanded as:

$$\begin{aligned} \sum_{\mu\nu} P_{\mu\nu}^{\Delta} (V_{\mu\nu}^{sol})^x &= - \sum_{\mu\nu} P_{\mu\nu}^{\Delta} \sum_i \left\langle \mu \left| \frac{\mathbf{q}}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle^x \\ &= - \sum_{\mu\nu} P_{\mu\nu}^{\Delta} \sum_i \left\langle \mu \left| \frac{1}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle^x \mathbf{q} - \sum_{\mu\nu} P_{\mu\nu}^{\Delta} \sum_i \left\langle \mu \left| \frac{1}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \nu \right\rangle \mathbf{q}^x \end{aligned} \quad (14)$$

where i runs over all surface tesserae.

The derivative of apparent surface charges \mathbf{q}^x can be converted back to potential derivative by using the CPCM linear response equation.

$$\begin{aligned}
& \sum_{\mu\nu} P_{\mu\nu}^{\Delta} \sum_i \left\langle \mu \left| \frac{1}{\|\mathbf{r}_e - \mathbf{r}_i\|} \right| \nu \right\rangle \mathbf{q}^x \\
&= \sum_{\mu\nu} P_{\mu\nu}^{\Delta} \sum_i \left\langle \mu \left| \frac{1}{\|\mathbf{r}_e - \mathbf{r}_i\|} \right| \nu \right\rangle \left(-\frac{\varepsilon-1}{\varepsilon} \mathbf{C}^{-1} \mathbf{V} \right)^x \\
&= -\left(\frac{\varepsilon-1}{\varepsilon} \right) \sum_{\mu\nu} P_{\mu\nu}^{\Delta} \sum_i \left\langle \mu \left| \frac{1}{\|\mathbf{r}_e - \mathbf{r}_i\|} \right| \nu \right\rangle \left[(\mathbf{C}^{-1})^x \mathbf{V} + \mathbf{C}^{-1} \mathbf{V}^x \right] \\
&= \left(\frac{\varepsilon}{\varepsilon-1} \right) \left[-\left(\frac{\varepsilon-1}{\varepsilon} \right) \mathbf{C}^{-1} \sum_{\mu\nu} P_{\mu\nu}^{\Delta} \sum_i \left\langle \mu \left| \frac{1}{\|\mathbf{r}_e - \mathbf{r}_i\|} \right| \nu \right\rangle \right] \mathbf{C}^x \left[-\left(\frac{\varepsilon-1}{\varepsilon} \right) \mathbf{C}^{-1} \mathbf{V} \right] \\
&\quad + \left[-\left(\frac{\varepsilon-1}{\varepsilon} \right) \mathbf{C}^{-1} \sum_{\mu\nu} P_{\mu\nu}^{\Delta} \sum_i \left\langle \mu \left| \frac{1}{\|\mathbf{r}_e - \mathbf{r}_i\|} \right| \nu \right\rangle \right] (\mathbf{V}^{nuc})^x \\
&\quad + \left[-\left(\frac{\varepsilon-1}{\varepsilon} \right) \mathbf{C}^{-1} \sum_{\mu\nu} P_{\mu\nu}^{\Delta} \sum_i \left\langle \mu \left| \frac{1}{\|\mathbf{r}_e - \mathbf{r}_i\|} \right| \nu \right\rangle \right] (\mathbf{V}^{ele})^x \\
&= \left(\frac{\varepsilon}{\varepsilon-1} \right) \mathbf{q}_{\mu\nu}^{\Delta} \mathbf{C}^x \mathbf{q} + \mathbf{q}_{\mu\nu}^{\Delta} (\mathbf{V}^{nuc})^x - \sum_{\rho\sigma} P_{\rho\sigma} \sum_i \left\langle \rho \left| \frac{1}{\|\mathbf{r}_e - \mathbf{r}_i\|} \right| \sigma \right\rangle^x \mathbf{q}_{\mu\nu}^{\Delta}
\end{aligned} \tag{15}$$

where $\mathbf{q}_{\mu\nu}$ represents PCM induced charge due to the TDDFT excitation difference density:

$$\mathbf{q}_{\mu\nu} = \left(\frac{\varepsilon-1}{\varepsilon} \right) \mathbf{C}^{-1} \mathbf{V}_{\mu\nu} = -\left(\frac{\varepsilon-1}{\varepsilon} \right) \mathbf{C}^{-1} \sum_{\mu\nu} P_{\mu\nu}^{\Delta} \sum_i \left\langle \mu \left| \frac{1}{\|\mathbf{r}_e - \mathbf{r}_i\|} \right| \nu \right\rangle \tag{16}$$

Therefore, Eq (14) becomes:

$$\begin{aligned}
\sum_{\mu\nu} P_{\mu\nu}^{\Delta} (V_{\mu\nu}^{sol})^x &= -\sum_{\mu\nu} P_{\mu\nu}^{\Delta} \sum_i \left\langle \mu \left| \frac{1}{\|\mathbf{r}_e - \mathbf{r}_i\|} \right| \nu \right\rangle^x \mathbf{q} - \sum_{\rho\sigma} P_{\rho\sigma} \sum_i \left\langle \rho \left| \frac{1}{\|\mathbf{r}_e - \mathbf{r}_i\|} \right| \sigma \right\rangle^x q_{\mu\nu}^{\Delta} \\
&\quad + \left(\frac{\varepsilon}{\varepsilon-1} \right) q_{\mu\nu}^{\Delta} \mathbf{C}^x \mathbf{q} + q_{\mu\nu}^{\Delta} (\mathbf{V}^{nuc})^x
\end{aligned} \tag{17}$$

The first term of Eq (17) represents forces imposed on PCM induced charge \mathbf{q} due to the electrostatic field created by solute electronic difference density. The second and the

fourth term represents forces imposed on PCM induced electronic difference charge $\mathbf{q}_{\mu\nu}$ due to the electrostatic field created by solute electrons and nuclei. The third term represents the forces imposed on the induced charge q in the electrostatic field created by PCM induced electronic difference charge.

The seventh term in Eq (13) involves solvation potential energy $E_{a,jb}^{sol}$ derivative and can be expanded as:

$$\begin{aligned}
& \sum_{\mu\nu\rho\sigma} (X+Y)_{\mu\nu} (X+Y)_{\rho\sigma} \left(E_{\mu\nu,\rho\sigma}^{sol} \right)^x \\
&= \sum_{\mu\nu\rho\sigma} (X+Y)_{\mu\nu} (X+Y)_{\rho\sigma} \left[V_{\mu\nu} \left(-\frac{\varepsilon-1}{\varepsilon} \right) \mathbf{C}^{-1} V_{\rho\sigma} \right]^x \\
&= 2 \sum_{\mu\nu\rho\sigma} (X+Y)_{\mu\nu} (X+Y)_{\rho\sigma} \left[V_{\mu\nu} \left(-\frac{\varepsilon-1}{\varepsilon} \right) \mathbf{C}^{-1} V_{\rho\sigma}^x \right] \\
&\quad + \sum_{\mu\nu\rho\sigma} (X+Y)_{\mu\nu} (X+Y)_{\rho\sigma} \left[V_{\mu\nu} \left(-\frac{\varepsilon-1}{\varepsilon} \right) (\mathbf{C}^{-1})^x V_{\rho\sigma} \right] \tag{18} \\
&= -2 \sum_{\rho\sigma} \left[\left(\frac{\varepsilon-1}{\varepsilon} \right) \mathbf{C}^{-1} \sum_{\mu\nu} (X+Y)_{\mu\nu} V_{\mu\nu} \right] (X+Y)_{\rho\sigma} \sum_i \left\langle \rho \left| \frac{1}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \sigma \right\rangle^x \\
&\quad + \left(\frac{\varepsilon}{\varepsilon-1} \right) \left[\left(\frac{\varepsilon-1}{\varepsilon} \right) \mathbf{C}^{-1} \sum_{\mu\nu} (X+Y)_{\mu\nu} V_{\mu\nu} \right] \mathbf{C}^x \left[\left(\frac{\varepsilon-1}{\varepsilon} \right) \mathbf{C}^{-1} \sum_{\rho\sigma} (X+Y)_{\rho\sigma} V_{\rho\sigma} \right] \\
&= 2 \sum_{\rho\sigma} (X+Y)_{\rho\sigma} \sum_i \left\langle \rho \left| \frac{1}{|\mathbf{r}_e - \mathbf{r}_i|} \right| \sigma \right\rangle^x q_{XY} + \left(\frac{\varepsilon}{\varepsilon-1} \right) q_{XY} \mathbf{C}^x q_{XY}
\end{aligned}$$

where q_{XY} is a set of CPCM surface charges induced by the TDDFT transition state density $(X+Y)$:

$$q_{XY} = \left(\frac{\varepsilon-1}{\varepsilon} \right) \mathbf{C}^{-1} \sum_{\mu\nu} (X+Y)_{\mu\nu} V_{\mu\nu} \tag{19}$$

Eq (18) represents the forces imposed on PCM induced electronic charge q_{XY} due to the electrostatic field created by solute electronic transition state density (X+Y), and the forces between q_{XY} .

The expression of Eq (13) is added to the ground state DFT-CPCM gradient to give the total free energy gradient expression. The evaluation of TDDFT-CPCM gradient requires \mathbf{C}^x , which involves the derivatives of the areas and coordinates of the tesserae with respect to the atomic coordinate x . In FIXPVA, the tesserae areas are smooth functions of their distances to neighboring spheres, so rigorously continuous and smooth potential energy surfaces, as well as exact analytical gradients, can be obtained.

2.3 Implementation and Tests

Accuracy of the gradients

TDDFT/CPCM gradients have been implemented in the GAMESS^[18] package based on the TDDFT and TDDFT/CPCM code implemented previously by Chiba et al^[19]. The CPCM code was previously implemented by Li and Jensen^[20] on the basis of the IEFPCM program originally implemented by Mennucci, Cancès, Pomelli and Tomasi.^[16, 21]

The analytic and numerical gradients were calculated for acetate embedded in bulk water. The numerical gradients were determined via double displacement with a step size of 0.001 Bohrs in each step for translational motions of the acetate atoms and water molecules. The acetate molecule was described with B3LYP/6-311G(d,p); the bulk water was described with CPCM ($\epsilon = 78.39$). The geometry of acetate was optimized with the

gas phase B3LYP/6-31G* method. In the B3LYP calculation, 96 radial and 302 Lebedev angular grid points were used. In the CPCM calculations, spheres with radii of 0, 2.124, 2.016 and 1.908 were used for H, C, N and O atoms, respectively, to define the molecular cavity; no additional spheres were used. Using zero radii for H atoms means that they do not contribute to form the surface. The tessellation scheme FIXPVA was used with 240 initial tesserae per sphere. The induced surface charges were determined by a semi-iterative DIIS procedure^[21b, 22] with no charge renormalization. Only the electrostatic interaction was considered; cavitation, dispersion and repulsion terms were not considered.

The agreement between the numerical and analytic gradients of TDDFT-CPCM is the same as the agreement in gas phase TDDFT, with most of the differences lying in between 10^{-6} and 10^{-5} au. The maximum unsigned difference is 29×10^{-6} au. Since the errors should be mainly numerical, the analytic gradients could be more accurate. It is noted that such accurate gradients are partially due to the use of the FIXPVA tessellation scheme developed by the Li's group.

PNA (p-nitroaniline)

Tomasi et al derived and implemented a TDDF/IEFPCM gradient and used this method to study the excited state behaviors of PNA in solution. The excited state geometries and properties of PNA are significantly influenced by solvent effects. In the current study, PNA was selected to test TDDFT/CPCM gradient implementation by comparing the optimized geometry parameters with Tomasi's result.

PNA demonstrates a strong absorption band in the near ultraviolet to visible spectral region based on intermolecular charge transfer (ICT) from $-\text{NH}_2$ to $-\text{NO}_2$. The experimental value of this excitation energy *in vacuo* is 4.24 eV ^[23] and it shows a red shift in solution to 3.85 eV in cyclohexane and 3.41 eV in acetonitrile. ^[24]

The ground state geometry of PNA was optimized by B3LYP/6-311G(d,p) methods in the gas phase and B3LYP/CPCM/6-311G(d,p) in cyclohexane and acetonitrile solution with dielectric constant 2.02 for cyclohexane and 36.64 for acetonitrile. The optimized geometry was used to perform a single point energy calculation by using TD-B3LYP/6-311G(d,p). Bulk water was described by nonequilibrium CPCM. The calculated excitation energy is, respectively, 4.06 eV in vacuum, 3.78 eV in cyclohexane and 3.59 eV in acetonitrile, which is comparable with experimental results and the calculated value by Tomasi et al (4.07 eV in vacuum, 3.83 eV in cyclohexane and 3.63 eV in acetonitrile).

The excited state geometry was optimized by TD-B3LYP/CPCM/6-311G(d,p) in cyclohexane and acetonitrile. The hessian calculation was performed to confirm the minimum geometry obtained. The geometry parameters for ground and excited state are listed in Table 2.1. The data in brackets refer to literature^[6a, 15]. The data from literature and current calculation are comparable. The difference is due to sphere radii difference of the two methods, in addition, B3LYP is different in Gaussian and GAMESS.

In ground state, the molecule is essentially planar with amine a little twisted. The twisting decreases with increasing solvent polarity. The dihedral angle decreases from 14.5 degrees in cyclohexane to 5.5 degrees in acetonitrile.

In excited state, nitro twisted structure was observed in cyclohexane and acetonitrile. The dihedral angle is almost the same in the two solvents with 71.8 degrees in cyclohexane and 71.5 degrees in acetonitrile. The bond lengths are not symmetric due to nitro O closer to a certain benzene region. The similar geometry properties of PNA in excited state demonstrate that current implementation is accurate.

Acetone

Acetone has been widely studied experimentally and theoretically. It serves in theoretical chemistry as a test molecule for the effect of solvent on the $n \rightarrow \pi^*$ transition. In our current work, the TDDFT-CPCM method was used to study the first excited state of acetone in aqueous solution through its geometry.

The transition energy from the oxygen non-bonding n orbital to the carbonyl anti-bonding π^* orbital is 4.38 eV in vacuum, measured using electron-impact spectroscopy^[25]. A ~ 0.2 eV blue shift in aqueous solution was experimentally observed^[26].

Acetone was optimized by B3LYP/6-311++G(2df,p) with solvent effects described by CPCM. Previous CPMD (Car-Parrinello Molecular dynamics) simulation shows two water molecules hydrogen bonded to the carbonyl oxygen in aqueous solution^[27]. The acetone-two-water cluster was selected as a model molecule and optimized by CPCM/B3LYP/6-311++G(2df,p).

By using the optimized ground state geometry, single point energy calculation was performed by TD-B3LYP/6-311++G(2df,p) method with nonequilibrium CPCM describing the solvation energy. The calculated $n \rightarrow \pi^*$ transition energy is, respectively,

4.43 eV in gas phase for acetone, 4.55 eV in aqueous solution with 0.12 eV blue shift for acetone, and 4.76 eV in aqueous solution with 0.33 eV blue shift for acetone-water₂. It is clear that single point CPCM calculation method on acetone underestimates the solvation effect and on acetone two water cluster overestimates the solvation effect. Yoo et al reproduce the excitation energy blue shift by a TDDFT/IEFPCM simulation.^[28]

The geometry of the π^* state of acetone and acetone-two-water cluster was optimized using CPCM/TD-B3LYP/6-311++G(2df,p). The optimized geometry parameters of the ground and excited states are listed in Table 2.2. The gas phase ground state experimental^[29] geometries are listed for comparison. The carbonyl bond length and C-C-C angle are in good agreement with experimental values, while the C-C bond length has a 0.08 Å deviation in the ground state gas phase. In the ground state, the C-O bond length increases and the corresponding C-C bond length decreases as a result of the aqueous solution effect, in agreement with literature.^[30] In the excited state, the C-O bond length is almost unchanged, while the C-C bond length increases. The distance between the water hydrogen and carbonyl oxygen is increased from the ground state to the excited state. The two water molecules are almost in the same plane with acetone in ground state while in the excited state the two waters are perpendicular to the acetone plane (Fig. 2.1). All of the above geometry changes can be explained by the shape and electronic density properties of the n orbital and π^* orbital.

Photoactive yellow protein (PYP)

Photoactive yellow protein (PYP, PDB: 2PHY), a photoreceptor in the bacterium *Ectothiorhodospira halophila*,^[31] demonstrates a maximum absorbance at 446 nm (2.78 eV) resulting in its bright yellow color. It is a small soluble protein containing 125 amino acids and a chromophore, a deprotonated p-coumaric acid covalently bonded to Cys69 through a thioester linkage. The chromophore O forms a hydrogen bond with Glu46 OH and Tyr42 OH. The ground state crystal structure of PYP was determined by Gezoff et al in 1995^[32].

Upon absorbing blue light, the PYP undergoes a series of intermediate states involving a *trans*-to-*cis* isomerization, then protonation of the chromophore, and finally a reversion back to the ground state^[33]. An early intermediate involves flipping the thioester group, reducing the perturbation of the protein interior, and twisting the double bond of the chromophore.^[34] Excited state simulations of a chromophore analogue in water demonstrate that the excited state decay in aqueous solution involves a single-bond (SB) twisted S1 minimum, in which phenyl-adjacent single bond is twisted, and a double bond twisted S1 minimum^{[35],[36]}. The predominant decay channel is the single bond twisted intermediate which demonstrates a lower barrier in aqueous solution.

In current study, a 22-atom model molecule (Figure 2.2) was used to study the intrinsic behavior of the PYP chromophore. A 43-atom model molecule with two low-barrier hydrogen bonds^[37] was also considered to try to involve the protein environment. The ground state geometry was optimized by the CPCM/B3LYP/ACCD method. The first excited state geometry was optimized with the CPCM/TD-B3LYP/ACCD method. In the CPCM calculation, dielectric constants of 78.39 and 5.00 were used to represent

aqueous and protein solvation. The optimized geometry parameters and relaxation energies are listed in Table 3.

For double bond twisted pyp 43-atom and 22-atom model molecules, the double bond is lengthened by 0.1 Å, giving it significant single bond character while the two single bonds adjacent to that double bond have shortened bond length with some double bond character. The distance between chromophore O and Glu 46 O as well as Tyr 42 slightly decreases. For single bond twisted pyp 43-atom and 22-atom model molecules, the double bonds C=O and C=C as well as the single bond adjacent to the benzene ring are lengthened while the single bond adjacent to S is shortened. Geometry is very similar for the model molecules with dielectric constant 78.39 and 5.

The calculated vertical excitation energies are, respectively, 3.03 and 3.17 eV for the 22-atom and 43-atom model molecules in aqueous solution. When solvent relaxation is considered, the excitation energy is lowered to 2.75 and 2.97 eV. It is apparent, therefore, that solvent relaxation can stabilize the electronic excited state. When both solvent and solute geometry relaxation effects are considered, the excitation energies are further lowered to 2.67 eV and 2.76 eV. The dielectric constant effect is small for the planar chromophore with an increase of 0.08 eV for the 22-atom model molecule, and no dielectric effect for the 43-atom model molecule.

For the 22-atom and 43-atom single bond twisted model molecules, the excitation energies for solvent and solute relaxed geometry are 1.39 eV and 1.79 (1.74) eV. 1.79 eV is found with a rotating benzene ring while 1.74 eV is found when the the opposite group is rotated. The dielectric effect is also more apparent, with a ~0.2 eV decrease for the 22-

atom model molecule and a ~ 0.1 eV decrease for the 43-atom model molecule with the dielectric constant decreased from 78.39 to 5.

For 22-atom and 43-atom double bond twisted model molecules, a low excitation energy for the relaxed geometry is obtained with 0.62 eV for 22-atom model molecule and 0.23 eV for 43-atom model molecule in aqueous solution, compared with planar and single bond twisted geometries. The excitation energy is further lowered to 0.58 eV and 0.15 eV in media with dielectric constant 5. Only the group opposite with benzene ring is rotated for the 43-atom model molecule. It is reported that hydrogen bonds between chromophore oxygen and Glu46, Tyr42 hydroxy do not seem to be perturbed for the double bond twisted geometry.

2.4 Conclusion

In summary, exact and continuous TDDFT/CPCM analytical gradients were obtained by using FIXPVA tessellation method. TDDFT/CPCM/ACCD geometry optimization demonstrates that the double bond twisted chromophore geometry of PYP is the most likely excited state decay channel in aqueous solution and in the protein.

References

- [1] E. Runge, E. K. U. Gross, *Physical Review Letters* **1984**, 52, 997.
- [2] aM. E. Casida, in *Recent Advances in Density Functional Methods* (Ed.: D. P. Chong), World Scientific, Singapore, **1995**, p. 155; bM. E. Casida, C. Jamorski, K. C. Casida, D. R. Salahub, *Journal of Chemical Physics* **1998**, 108, 4439-4449.
- [3] A. Dreuw, M. Head-Gordon, *Chemical Reviews* **2005**, 105, 4009-4037.
- [4] T. Ziegler, M. Seth, M. Krykunov, J. Autschbach, F. Wang, *The Journal of Chemical Physics* **2009**, 130, -.
- [5] aC. Van Caillie, R. D. Amos, *Chemical Physics Letters* **1999**, 308, 249-255; bC. Van Caillie, R. D. Amos, *Chemical Physics Letters* **2000**, 317, 159-164.
- [6] aF. Furche, R. Ahlrichs, *Journal of Chemical Physics* **2002**, 117, 7433-7447; bF. Furche, R. Ahlrichs, *Journal of Chemical Physics* **2004**, 121, 12772-12773.
- [7] N. C. Handy, H. F. Schaefer III, *The Journal of Chemical Physics* **1984**, 81, 5031-5033.
- [8] S. Miertus, E. Scrocco, J. Tomasi, *Chemical Physics* **1981**, 55, 117-129.
- [9] E. Cancès, B. Mennucci, J. Tomasi, *The Journal of Chemical Physics* **1997**, 107, 3032-3041.
- [10] V. Barone, M. Cossi, *The Journal of Physical Chemistry A* **1998**, 102, 1995-2001.
- [11] P. Su, H. Li, *The Journal of Chemical Physics* **2009**, 130, -.
- [12] aA. W. Lange, J. M. Herbert, *The Journal of Physical Chemistry Letters* **2009**, 1, 556-561; bG. Scalmani, M. J. Frisch, *The Journal of Chemical Physics* **2010**, 132, -.
- [13] aA. Öhrn, G. Karlström, *Molecular Physics* **2006**, 104, 3087-3099; bLin, J. Gao, *Journal of Chemical Theory and Computation* **2007**, 3, 1484-1493; cJ. Kongsted, A. Osted, K. V. Mikkelsen, O. Christiansen, *The Journal of Chemical Physics* **2003**, 118, 1620-1633; dL. Jensen, P. T. van Duijnen, J. G. Snijders, *The Journal of Chemical Physics* **2003**, 119, 3800-3809; eS. Yoo, F. Zahariev, S. Sok, M. S. Gordon, *The Journal of Chemical Physics* **2008**, 129, -.
- [14] M. Cossi, V. Barone, *Journal of Chemical Physics* **2001**, 115, 4708-4717.
- [15] G. Scalmani, M. J. Frisch, B. Mennucci, J. Tomasi, R. Cammi, V. Barone, *Journal of Chemical Physics* **2006**, 124, 15.
- [16] E. Cancès, B. Mennucci, J. Tomasi, *Journal of Chemical Physics* **1997**, 107, 3032-3041.
- [17] aB. Mennucci, R. Cammi, J. Tomasi, *Journal of Chemical Physics* **1998**, 109, 2798-2807; bM. Cossi, V. Barone, *Journal of Chemical Physics* **2000**, 112, 2427-2435.
- [18] aM. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *Journal of Computational Chemistry* **1993**, 14, 1347-1363; bM. S. Gordon, M. W. Schmidt, *Advances in electronic structure theory: GAMESS a decade later*, Elsevier, Amsterdam, **2005**.
- [19] aM. Chiba, T. Tsuneda, K. Hirao, *Journal of Chemical Physics* **2006**, 124, -; bM. Chiba, T. Tsuneda, K. Hirao, *Chemical Physics Letters* **2006**, 420, 391-396.
- [20] H. Li, J. H. Jensen, *Journal of Computational Chemistry* **2004**, 25, 1449-1462.

- [21] aJ. Tomasi, B. Mennucci, E. Cancès, *Journal of Molecular Structure-Theochem* **1999**, 464, 211-226; bH. Li, C. S. Pomelli, J. H. Jensen, *Theoretical Chemistry Accounts* **2003**, 109, 71-84.
- [22] C. S. Pomelli, J. Tomasi, V. Barone, *Theoretical Chemistry Accounts* **2001**, 105, 446-451.
- [23] S. Millefiori, G. Favini, A. Millefiori, D. Grasso, *Spectrochimica Acta Part A: Molecular Spectroscopy* **1977**, 33, 21-27.
- [24] aS. A. Kovalenko, R. Schanz, V. M. Farztdinov, H. Hennig, N. P. Ernsting, *Chemical Physics Letters* **2000**, 323, 312-322; bV. M. Farztdinov, R. Schanz, S. A. Kovalenko, N. P. Ernsting, *The Journal of Physical Chemistry A* **2000**, 104, 11486-11496.
- [25] K. N. Walzl, C. F. Koerting, A. Kuppermann, *The Journal of Chemical Physics* **1987**, 87, 3796-3803.
- [26] N. S. Bayliss, E. G. McRae, *The Journal of Physical Chemistry* **2002**, 58, 1006-1011.
- [27] aM. Pavone, O. Crescenzi, G. Morelli, N. Rega, V. Barone, *Theoretical Chemistry Accounts* **2006**, 116, 456-461; bO. Crescenzi, M. Pavone, F. De Angelis, V. Barone, *Journal of Physical Chemistry B* **2005**, 109, 445-453.
- [28] S. Yoo, F. Zahariev, S. Sok, M. S. Gordon, *The Journal of Chemical Physics* **2008**, 129, 144112.
- [29] R. L. Hilderbrandt, A. L. Andreassen, S. H. Bauer, *The Journal of Physical Chemistry* **2002**, 74, 1586-1592.
- [30] F. Aquilante, M. Cossi, O. Crescenzi, G. Scalmani, V. Barone, *Molecular Physics* **2003**, 101, 1945-1953.
- [31] M. Baca, G. E. O. Borgstahl, M. Boissinot, P. M. Burke, D. R. Williams, K. A. Slater, E. D. Getzoff, *Biochemistry* **2002**, 33, 14369-14377.
- [32] G. E. O. Borgstahl, D. R. Williams, E. D. Getzoff, *Biochemistry* **2002**, 34, 6278-6287.
- [33] M. J. Thompson, D. Bashford, L. Noodleman, E. D. Getzoff, *Journal of the American Chemical Society* **2003**, 125, 8186-8194.
- [34] U. K. Genick, S. M. Soltis, P. Kuhn, I. L. Canestrelli, E. D. Getzoff, *Nature* **1998**, 392, 206-209.
- [35] M. Boggio-Pasqua, M. A. Robb, G. Groenhof, *Journal of the American Chemical Society* **2009**, 131, 13580-13581.
- [36] A. M. Virshup, C. Punwong, T. V. Pogorelov, B. A. Lindquist, C. Ko, T. J. Martiñánez, *The Journal of Physical Chemistry B* **2008**, 113, 3280-3291.
- [37] S. Yamaguchi, H. Kamikubo, K. Kurihara, R. Kuroki, N. Niimura, N. Shimizu, Y. Yamazaki, M. Kataoka, *Proceedings of the National Academy of Sciences of the United States of America* **2009**, 106, 440-444.

Table 2.1 Ground state and excited state geometry parameters of PNA, respectively, optimized by B3LYP/6-311G(d,p) and TD- B3LYP/6-311G(d,p) in gas phase, cyclohexane and acetonitrile; length with Å and angle with degree.

	Ground State				Excited State
	Gas Phase	Cyclohexane	Acetonitrile	Cyclohexane	Acetonitrile
R(C ₂ N ₁₁)	1.378 (1.367)	1.367 (1.370)	1.353 (1.357)	1.344 (1.344)	1.334 (1.333)
R(C ₂ C _{5/6})	1.409 (1.409)	1.414 (1.412)	1.420 (1.417)	1.428 (1.427)	1.432 (1.432)
R(C _{5/6} C _{4/3})	1.384 (1.382)	1.381 (1.381)	1.378 (1.378)	1.374/1.367 (1.367)	1.374/1.366 (1.367)
R(C _{4/3} C ₁)	1.395 (1.394)	1.397 (1.396)	1.402 (1.400)	1.409/1.414 (1.414)	1.408/1.413 (1.414)
R(C ₁ N ₁₂)	1.462 (1.462)	1.455 (1.456)	1.441 (1.442)	1.428 (1.428)	1.430 (1.429)
R(N ₁₂ O)	1.229 (1.227)	1.231 (1.230)	1.237 (1.235)	1.306 (1.305)	1.309 (1.307)
H ₁₄ N ₁₁ -C ₂ C ₅	19.3 (19.2)	14.5 (17.2)	5.5 (11.2)	-0.42 (0.0)	-0.4 (0.0)
O ₁₆ N ₁₂ -C ₁ C ₄	0.7 (0.0)	1.0 (0.0)	2.0 (0.0)	71.8 (71.8)	71.5 (72.6)

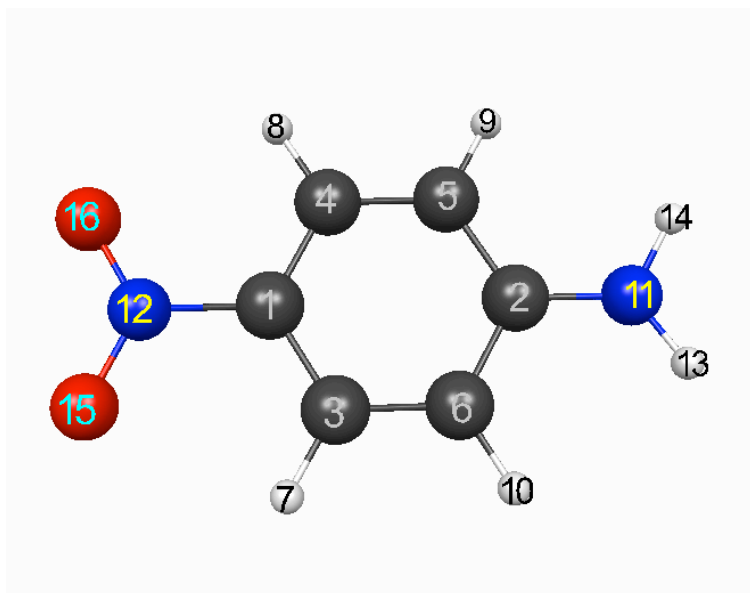


Table 2.2 Geometrical parameters of acetone and acetone+two water cluster geometry optimized in vacuo and in aqueous solution by B3LYP/6-311++G(2df,p) and TD-B3LYP/6-311G+(2df,p); distances in Å, angle in degrees

	acetone	acetone+two water cluster				acetone	
	Exp./vacuum	DFT	TDDFT		DFT	TDDFT	
			CPCM	Gas	CPCM	Gas	CPCM
C=O	1.210	1.227	1.311	1.210	1.219	1.305	1.304
C-C	1.507	1.500	1.513	1.515	1.507	1.506	1.514
C-C=O	121.7	121.5	114.7	121.7	121.6	114.8	114.7
C-C-C	116.7	117.1	120.0	116.6	116.7	120.0	119.2

Acetone+two water cluster		
	DFT/CPCM	TDDFT/CPCM
O...H _{water1}	1.899	2.015
O...H _{water2}	1.899	1.984
O...O _{water1}	2.873	2.983
O...O _{water2}	2.873	2.952
C=O...H _{water1}	125.4	126.7
C=O...H _{water2}	127.0	119.6
O _{actone} ...H-O _{water1}	177.1	174.7
O _{actone} ...H-O _{water2}	178.5	174.9
CC=O...H _{water1}	-3.59	69.86
CC=O...H _{water2}	-2.16	109.7

Figure 2.1 Acetone and acetone+two water cluster optimized in aqueous solution with CPCM/B3LYP/6-311++G(2df,p) and CPCM/TD-B3LYP/6-311G+(2df,p) methods.

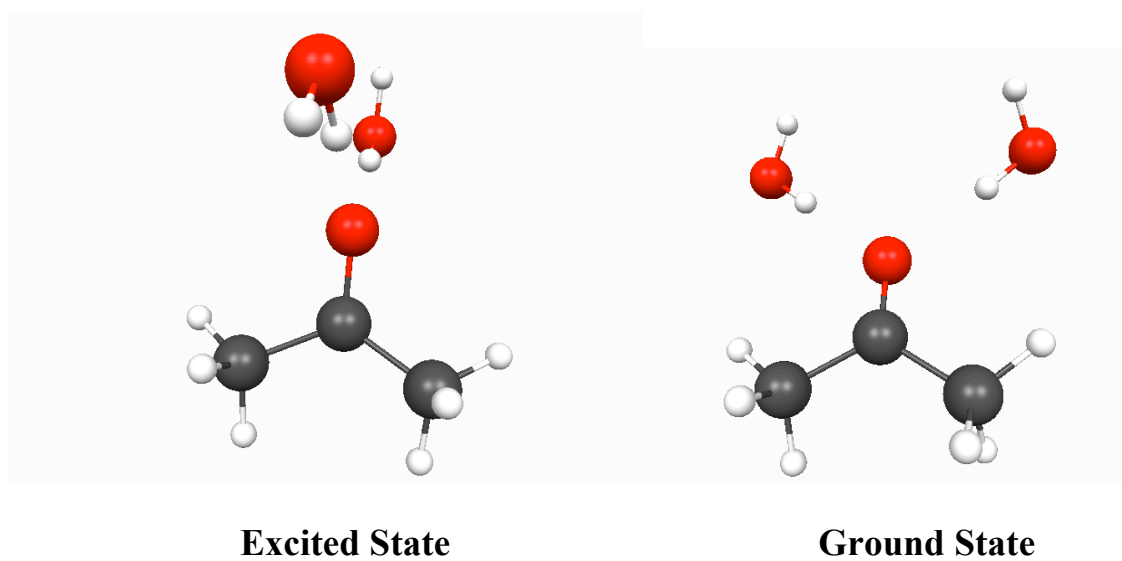


Table 2.3 Geometrical parameters of 22-atom and 43-atom model molecule of PYP

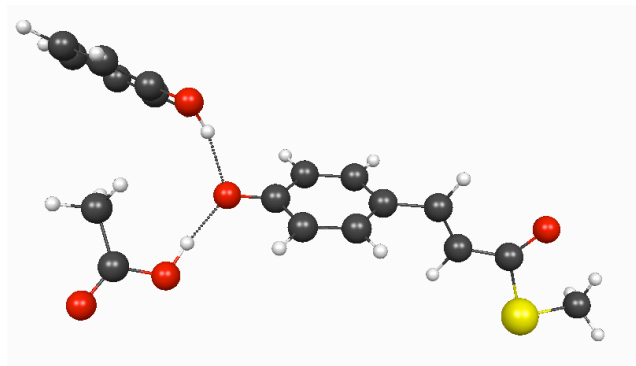
optimized in aqueous solution by TD-B3LYP/ACCD; distances in Å, angle in degrees.

	C=O	C=C	C- CO	phenylC- C	O...O _{Glu}	O...O _{Tyr}	C ₁₀ C ₁₁ C ₁₂ C ₁₃
2PHY	1.237	1.304	1.509	1.488	2.56	2.52	-2.4
22-atom/ground	1.234	1.374	1.447	1.432			0.1
excited	1.255	1.401	1.439	1.445			-0.9
single bond	1.266	1.464	1.418	1.480			-89.4
twisted							
Double bond	1.235	1.464	1.437	1.408			
twisted							
43-atom/ground	1.231	1.365	1.459	1.445	2.56	2.616	
excited	1.253	1.406	1.435	1.436	2.629	2.638	
single bond	1.265	1.397	1.419	1.476	2.689	2.786	
twisted							
Double bond	1.236	1.465	1.435	1.406	2.549	2.588	
twisted							

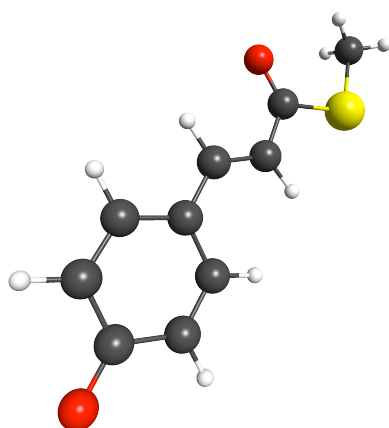
Table 2.4 Relaxation energies (eV) of 22-atom and 43-atom model molecules of PYP optimized in aqueous solution by TD-B3LYP/ACCD.

	Single bond twisted	Double bond twisted	flat	Vertical excitation energy
22-atom/eps=78.39	-0.61	-0.68	-0.33	3.028
22-atom/eps=5	-0.78	-0.72	-0.23	3.009
43-atom/eps=78.39	-0.28	-0.89	-0.34	3.178
43-atom/eps=5	-0.75	-12.4		3.172

Figure 2.2 CPCM/TD-B3LYP/ACCD optimized 22-atom and 43-atom model molecules of PYP in aqueous solution.



43-atom model



22-atom model

Chapter 3

Protonation of His120 in Cu_A Azurin: A Quantum Chemical Study

Abstract

The C-terminal His120 in Cu_A azurin has the experimental pK_a value: 4.8±0.1. the quantum chemical method C-PCM/B3LYP/6-31G* is used to optimize 145-atom and 144-atom models extracted from the X-ray crystal structure 1CC3 for the acid and base-forms of His120. The calculated pK_a of His120 of 145-atom model is about 4.6 using a relative pK_a calculation method with 8.4 as the pK_a of reference 4-methylimidazole. The calculated pK_a for a 142-atom no-water model is -10.7, which is lower than the experimental pK_a value. Our research supports the hypothesis that a water molecule is coordinated with the copper of Cu_A. It is very possible that the pK_a of His120 is very sensitive to the model structure and is mainly affected by solvation effects.

3.1 Introduction

A binuclear redox active Cu_A center is responsible for inter- and intra-molecular electron transfer in cytochrome c oxidase (CcO),^[1] nitrous oxide reductase (N₂OR)^[2] and nitric oxide reductase (NOR).^[3] CcO utilizes four electrons and four protons to reduce O₂ to two H₂O; N₂OR uses two electrons and two protons to reduce nitrous oxide to nitrogen; ^[4] NOR uses two electrons and two protons to reduce nitric oxide to nitrous oxide. ^[5] Crystal structures of Cu_A proteins have been determined, for example, for *Thermus thermophilus* ba₃-type cytochrome c oxidase Cu_A domain (2CUA in PDB),^[6] Cu_A azurin (1CC3 in PDB)^[7] and purple Cu_A CyoA (1CYX in PDB).^[8] The crystal structures show that Cu_A center has a nearly planar diamond core Cu₂S₂ structure with two Cu ions bridged by two cysteine thiolates. Each Cu ion is further coordinated by a N¹(His) and either a S(Met) or a backbone carbonyl O of a Glu or a Gln. Experimental and computational methods have been used to study the geometric and electronic properties of Cu_A center. ^[9] It has been established that in the reduced state, the Cu_A consists of two Cu⁺ ions; in the oxidized state, it consists of two mixed valence Cu^{1.5+} ions.

The pH dependent behavior of the Cu_A activity may have a significant role in biological effects studied using various experimental methods.^[10] Shyamalava *et al* measured the pK_a of Cu_A domain (TtCu_A) in *Thermus thermophilus* ba₃-type cytochrome c oxidase and found that a pK_a value of ~5.5 is most likely due to the C-terminal histidine His157. ^[10a, 10d] Lu *et al* identified that the pK_a of the solvent exposed C-terminal His120 of Cu_A azurin is 4.8 at 277 K. Protonated His120 significantly decreases the electron transfer rate for Cu_A site by a factor of ~2000, and may serve to control the Cu_A activity.

[9i, 10c] Previous spectroscopic experimental and DFT calculations suggest that upon protonation, the imidazolium ring of His120 will de-coordinate the Cu_A site, and a nearby water will take its coordinating role. [9b, 9i]

In this work, a conductor like polarizable continuum model (CPCM^[11]) incorporated in the B3LYP^[12] as a reaction field is used to study the protonation of the His120 in Cu_A azurin. Model molecules consisting of 144 and 145 atoms were extracted from the X-ray structures 1CC3, and 85 and 86 atoms were optimized, while the rest of the atoms were fixed in their X-ray coordinates. For comparison, model molecules with and without a water molecule were both constructed and computed, and it was found that the Cu-coordination water has a significant effect on the Cu_A site structure and His120 pK_a. The calculated pK_a is 3.9 at 298 K, and is estimated as ~4.2 at 277 K, in good agreement with the experimental value 4.8 at 277 K.

3.2 Computational Method

A 141-atom base-form model was extracted from the chain A of the X-ray crystal structure 1CC3 for Cu_A azurin. [7] Hydrogen atoms were added to the PDB file using the WHAT IF web interface. [13] Based on the 141-atom model, a 142-atom acidic-form model, a 144-atom base-form (Figure 1) and a 145-atom acid-form model were constructed by manually adding protons to N^{si} of His120 to give the acid-form model and a H₂O molecule for the 144-atom and 145-atom model.

All the quantum chemical calculations were performed with the GAMESS package. [14] The main method used in this study is the unrestricted open shell B3LYP^[12] with 96 radial and 302 Lebedev angular grid points, and the standard 6-31G* basis set. In

the quantum chemical calculations of the Cu_A model, the oxidation states of the Cu ions are +1.5 and the electron spin multiplicity is 2.

Two CPCM methods were used. One is the commonly used homogeneous CPCM (Homo-CPCM^[11a, 11b]) in which only one dielectric constant can be defined and used to represent a homogeneous and isotropic solvent, such as bulk water. The other is a recently developed heterogeneous CPCM (Het-CPCM^[11c]) in which different local effective dielectric constants can be defined for different surface regions to represent a heterogeneous environment, such as an active site solvated by protein matrix and bulk water. Practically speaking, this is realized by defining different effective dielectric constants for different spheres used to form the solute cavity. In the current implementation, the CPCM boundary elements or tesserae on the same sphere have the same dielectric constant. In the Homo-CPCM method, dielectric constants of 78.39, 20, 10 and 4 were used. In the Het-CPCM calculations, a dielectric constant of 78.39 was used for the spheres associated with the solvent exposed atoms, and 20, 10 and 4 for spheres associated with the protein buried atoms.

In both the Homo-CPCM and Het-CPCM methods, spheres with radii of 0.0, 2.124, 2.016, 1.908, 2.52 and 2.76 Å, respectively, were used for H, C, N, O, S and Cu except for the atoms on the imidazole ring of His120, for which the United Atom model for Hartree-Fock (UAHF^[15]) radii 2.340, 2.286, 2.286, 2.016, 2.232, 2.016 Å, respectively, were used for C^β, C^γ, C^{δ2}, N^{δ1}, C^{ε1} and N^{ε2}. The FIXPVA surface tessellation scheme^[16] was used with 240 initial tesserae per sphere.

Geometry optimizations were performed with both the Homo- and Het-CPCM/U-B3LYP/6-31G*^[17] method. Some atoms were frozen to mimic the strain from the protein

(Figure 2). Single point energies were calculated with the following methods: gas phase B3LYP/6-31G*, gas phase B3LYP/mixed basis set, Homo-CPCM/B3LYP/mixed basis set, Homo-CPCM/B3LYP/cc-PVTZ^[18] for Cu-6-311++G(2df,p)^[19] for H, C, N, O and S, where mixed basis set represents aug-cc-PVTZ for Cu, 6-311++G(2df,p) for the atoms in imidazole ring of His120, S_{C112}, S_{C116}, backbone O_{His120}, backbone O_{Glu114} and the water, and 6-31G* for other atoms.

4-methyl-imidazolium (Imd) is used as the reference in the relative pK_a calculation of His120 in Cu_A azurin (H120). The standard free energy change for the following reaction



is approximated with the electronic energies (including nuclear repulsion and solvation energies) computed for the model molecules:

$$\Delta G \approx \Delta E^{ele} = \Delta E_{\text{H120}}^{ele} + \Delta E_{\text{Imd} \cdot \text{H}}^{ele} - \Delta E_{\text{H120} \cdot \text{H}}^{ele} - \Delta E_{\text{Imd}}^{ele} \quad (2)$$

The relative method is based on the assumption that the differences in the zero-point-energies, thermal energies and entropies of the model molecules make minor contributions to the relative pK_a. The free energy contributions due to the protein matrix and aqueous solvent are included in the CPCM solvation free energy.

The pK_a of the protein at T=298 K is computed by:

$$\text{pK}_a = 7.52 + \frac{\Delta G}{RT \ln 10} \quad (3)$$

where *R* is the gas constant and 7.52 is the experimental pK_a of 4-methyl-imidazolium at 298 K.^[20]

3.3 Results and Discussions

Model structure

The geometric parameters of the CPCM/U-B3LYP/6-31G* optimized model molecules and the corresponding values in the X-ray structure 1CC3 and those from an EXAFS measurement^[9i] are listed in Table 1 and shown in Figure 3. Cu_S is the Cu coordinated by S(Met123), Cu_O is the Cu coordinated by backbone O(Glu114), O_w and H_w are the oxygen and hydrogen atoms of the water molecule.

In the 144-atom model molecule the O_w-Cu_O distance is 4.93 Å, as compared to 5.11 Å in the X-ray structure 1CC3 chain A; the two H_w form hydrogen bonds with the backbone carbonyl O atoms of His120 (1.93 Å) and Cys112 (1.91 Å). In the 145-atom model molecule the O_w-Cu_O distance is 2.12 Å (Figure 3), as compared to 2.26 Å in a previous DFT calculation; ^[9i] the two H_w form hydrogen bonds with the backbone carbonyl O atoms of His120 (1.70 Å) and Cys112 (1.66 Å).

The Cu_S-N¹(His46) distances are 1.98 and 1.97 Å, respectively, and the Cu_O-N¹(His120) distances are 2.00 and 4.42 Å, respectively, in the 144-atom and 145-atom model molecules. EXAFS data show that in the high pH base-form there are two Cu-N(His) bonds both with a bond length of 1.92 Å, while in the low pH acid-form there is only one Cu-N(His) bond with a bond length of 1.95 Å. ^[9i] The calculated Cu-N bond lengths are longer than the EXAFS results. A previous DFT calculation gives Cu_S-N¹(His46) distances of 2.04 and 2.03 Å, respectively, in the 96-atom and 97-atom model molecules, and a Cu_O-N¹(His120) distance of 2.02 Å in the 96-atom model molecule. ^[9i]

The Cu_O-O(Glu114) distance is 2.10 and 2.00 Å, respectively, in the 144-atom and 145-atom model molecules, as compared to 2.17 Å in 1CC3 and 2.34 and 2.08 Å, respectively, from a previous DFT calculation with 96-atom and 97-atom gas phase

model molecules.^[9i] The shortening of the Cu_O-O(Glu114) distance on going from the base-form to the acid-form can be explained by the fact that Cu_O-O_W bond in the acid-form is weaker than the Cu_O-N^{δ⁺}(His120) bond, so the Cu_O-O(Glu114) bond becomes slightly stronger.

The Cu_S-S(Met123) distances are 2.72 and 2.73 Å, respectively, in the 144-atom and 145-atom model molecules, as compared to 2.98 Å in 1CC3 and 2.79 and 2.64 Å in a previous DFT calculation with 96-atom and 97-atom gas phase model molecules.^[9i]

The four Cu-S(Cys) distances are 2.29, 2.27, 2.27 and 2.26 Å in the 144-atom model molecule, and are 2.26, 2.26, 2.27 and 2.26 Å in the 145-atom model molecule. The average Cu-S(Cys) distances in the 144-atom and 145-atom model molecules are 2.27 and 2.26 Å, respectively, in excellent agreement with EXAFS results 2.27 and 2.26 Å for the base-form and acid-form (Table 1). The Cu-Cu distance is 2.39 and 2.34 Å, respectively, in the 144-atom and 145-atom model molecules, in excellent agreement with an EXAFS measurement that gives 2.40 and 2.36 Å, respectively, for the base-form and acid-form.^[9i] Therefore, upon protonation, the Cu-Cu distance is likely to shorten by 0.04~0.05 Å, consistent with the experimental observation that His120 binding can change the Cu-Cu distance.^[21] Since the Cu-S(Cys) distances are almost unchanged up on His120 protonation and de-coordination, the ~0.05 Å shortening of the Cu-Cu distance is due to the bond angle changes in the Cu₂S₂ core structure. The Cu-Cu distance is 2.42 Å in the crystal structure 1CC3 chain A, which corresponds to the base-form, but is in solid state instead of solution.

The Cu_S-Cu_O-N(His120) and Cu_O-Cu_S-N(His46) bond angles are 155° and 168°, respectively, in the crystal structure 1CC3. They are 149° and 150° in the optimized 144-

atom model molecule, and are 142° and 149° in the 145-atom model molecule, respectively. His120 imidazole ring and His46 imidazole ring are, respectively, above and below Cu_2S_2 plane. A Raman resonance spectroscopy analysis for *P. denitrificans* Cu_A domain (base-form at $\text{pH} \sim 10$) suggests that the two Cu-Cu-N angles are both $140 \pm 10^\circ$.^[22]

Therefore, the CPCM/B3LYP/6-31G* optimized 144- and 145-atom model molecules reasonably reproduce the high pH and low pH active site structures of Cu_A azurin in solution.

pK_a value

The experimental pK_a value is 4.8 ± 0.1 at 277 K.^[10c] The temperature dependence of this pK_a is not available. The solvent-exposed histidine pK_a in myoglobin^[23] and in mesophilic homologue of the peripheral subunit-binding domain family^[24] can decrease by 0.2~0.5 pH units on going from 277 K to 298 K. Therefore, the pK_a of His 120 can be estimated as 4.5 at 298 K.

The Homo-CPCM/B3LYP/6-31G* computed pK_a is 3.3 (Table 2). The calculated pK_a value based on the 6-31G*, 6-311++G(2df,p) and cc-pVTZ mixed basis set is 3.7. The calculated pK_a by Homo-CPCM in aqueous solution with basis set cc-pVTZ for Cu and 6-311++G(2df,p) is 3.9, the best prediction from this work.

Homo- and Het-CPCM/B3LYP/6-31G* geometry optimizations were performed with various dielectric constants. Using $\epsilon = 20, 10$ and 4 for atoms buried in protein and 78.39 for solvent exposed atoms, the calculated pK_a values are 2.5, 1.4 and -1.7 , respectively. Using $\epsilon = 20, 10$ and 4 in Homo-CPCM/B3LYP/6-31G*, the calculated pK_a

is 1.9, 0.0 and -5.4 , respectively. These data suggest that in crystallized Cu_A azurin the His120 should show a much lower pK_a value than 4.8, and should always be in the base-form. In other words, due to the solvation, the protonated His120 is relative stable.

The computed gas phase pK_a 8.2 deviates from the experimental value by 5 units based on the 631G* basis set and the higher pK_a value 10.40 was obtained from mixed basis set.

Influence of structural water on His120 pK_a

In the 144-atom base-form model, the water molecule forms two H-bonds. In the 145-atom acid-form model, the water molecule forms three H-bonds and one Cu-O coordinate covalent bond.

The 141-atom base-form model and the 142-atom acid-form model were optimized by Homo-CPCM/U-B3LYP/6-31G* to test the effect of the buried water for His120 pK_a . The optimized geometries of the 141-atom and 144-atom model molecules are similar, but the optimized geometry of the 142-atom model is very different from that of the 145-atom model. The calculated pK_a is -7.9 , deviating from the experimental value by ~ 12 pH units, corresponding to 16.9 kcal/mol.

The Cu_O-N(His120) distances are 3.58 Å and 4.42 Å, respectively, in the 142-atom and 145-atom model molecules. The angles formed by Cu_S-Cu_O-N¹(His120) and Cu_O-Cu_S-N¹(His46) in 142-atom model are 125° and 138°, as compared to 155° and 168° in 1CC3 and 142° and 149° in the 145-atom model. The 2.29 Å Cu-Cu distance in the 142-atom model molecule significantly deviates from the 2.42 Å in 1CC3.

Our calculation reveals that a structural water molecule plays a critical role for His120 pK_a through H-bonds and stabilizing rigid Cu_2S_2 diamond structure.

3.4 Conclusions

We conclude that the imidazoles are at an angle of about 30 degrees above and below the Cu_2S_2 plane. Protein matrix of the Cu_A center lowers the pK_a of His120 in Cu_A azurin since histine pK_a changes from 6.0 to 4.5 upon moving into Cu_A center.

References

- [1] aD. A. Mills, S. Xu, L. Geren, C. Hiser, L. Qin, M. A. Sharpe, J. McCracken, B. Durham, F. Millett, S. Ferguson-Miller, *Biochemistry* **2008**, *47*, 11499-11509; bO. Farver, Y. Chen, J. A. Fee, I. Pecht, *FEBS Letters* **2006**, *580*, 3417-3421; cL. Muresanu, P. Pristovsek, F. Lohr, O. Maneg, M. D. Mukrasch, H. Ruterjans, B. Ludwig, C. Lucke, *J. Biol. Chem.* **2006**, *281*, 14503-14513; dO. Farver, E. Grell, B. Ludwig, H. Michel, I. Pecht, *Biophysical Journal* **2006**, *90*, 2131-2137.
- [2] S. Dell'Acqua, S. R. Pauleta, E. Monzani, A. S. Pereira, L. Casella, J. J. G. Moura, I. Moura, *Biochemistry* **2008**, *47*, 10852-10862.
- [3] N. J. Watmough, S. J. Field, R. J. L. Hughes, D. J. Richardson, *Biochemical Society Transactions* **2009**, *037*, 392-399.
- [4] W. G. Zumft, P. M. H. Kroneck, K. P. Robert, in *Advances in Microbial Physiology*, Vol. Volume 52, Academic Press, **2006**, pp. 107-227.
- [5] U. Flock, P. Lachmann, J. Reimann, N. J. Watmough, P. fdelroth, *Journal of Inorganic Biochemistry* **2009**, *103*, 845-850.
- [6] P. A. Williams, N. J. Blackburn, D. Sanders, H. Bellamy, E. A. Stura, J. A. Fee, D. E. McRee, *Nat Struct Mol Biol* **1999**, *6*, 509-516.
- [7] H. Robinson, M. C. Ang, Y. G. Gao, M. T. Hay, Y. Lu, A. H. J. Wang, *Biochemistry* **1999**, *38*, 5677-5683.
- [8] M. Wilmanns, P. Lappalainen, M. Kelly, E. SauerEriksson, M. Saraste, *Proceedings of the National Academy of Sciences of the United States of America* **1995**, *92*, 11955-11959.
- [9] aS. I. Gorelsky, X. Xie, Y. Chen, J. A. Fee, E. I. Solomon, *Journal of the American Chemical Society* **2006**, *128*, 16452-16453; bH. J. Hwang, N. Nagraj, Y. Lu, *Inorganic Chemistry* **2006**, *45*, 102-107; cG. N. Ledesma, D. H. Murgida, H. K. Ly, H. Wackerbarth, J. Ulstrup, A. J. Costa, A. J. Vila, *Journal of the American Chemical Society* **2007**, *129*, 11884-+; dM. Wolpert, O. Maneg, B. Ludwig, P. Hellwig, **2004**, pp. 73-76; eA. X. Song, L. Z. Li, T. Yu, S. M. Chen, Z. X. Huang, *Protein Engineering* **2003**, *16*, 435-441; fB. Epel, C. S. Slutten, F. Neese, P. M. H. Kroneck, W. G. Zumft, I. Pecht, O. Farver, Y. Lu, D. Goldfarb, *Journal of the American Chemical Society* **2002**, *124*, 8152-8162; gS. D. George, M. Metz, R. K. Szilagyi, H. X. Wang, S. P. Cramer, Y. Lu, W. B. Tolman, B. Hedman, K. O. Hodgson, E. I. Solomon, *Journal of the American Chemical Society* **2001**, *123*, 5757-5767; hM. H. M. Olsson, U. Ryde, *Journal of the American Chemical Society* **2001**, *123*, 7866-7876; iX. Xie, S. I. Gorelsky, R. Sarangi, D. K. Garner, H. J. Hwang, K. O. Hodgson, B. Hedman, Y. Lu, E. I. Solomon, *J Am Chem Soc* **2008**, *130*, 5194-5205.
- [10] aN. J. Sanghamitra, S. Mazumdar, *Biochemistry* **2008**, *47*, 1309-1318; bS. Gupta, A. Warne, M. Saraste, S. Mazumdar, *Biochemistry* **2001**, *40*, 6180-6189; cH. J. Hwang, Y. Lu, *Proc Natl Acad Sci U S A* **2004**, *101*, 12842-12847; dA. Sujak, N. J. Sanghamitra, O. Maneg, B. Ludwig, S. Mazumdar, *Biophys J* **2007**, *93*, 2845-2851; eO. Farver, H. J. Hwang, Y. Lu, I. Pecht, *The Journal of Physical Chemistry B* **2007**, *111*, 6690-6694; fM. G. Savelieff, T. D. Wilson, Y. Elias, M. J. Nilges, D. K. Garner, Y. Lu, *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 7919-7924.

- [11] aV. Barone, M. Cossi, *The Journal of Physical Chemistry A* **1998**, *102*, 1995-2001; bH. Li, J. H. Jensen, *Journal of Computational Chemistry* **2004**, *25*, 1449-1462; cD. J. Si, H. Li, **Submitted**.
- [12] aR. H. Hertwig, W. Koch, *Chemical Physics Letters* **1997**, *268*, 345-351; bA. D. Becke, *Physical Review A* **1988**, *38*, 3098; cC. Lee, W. Yang, R. G. Parr, *Physical Review B* **1988**, *37*, 785.
- [13] R. Rodriguez, G. Chinae, N. Lopez, T. Pons, G. Vriend, *Bioinformatics* **1998**, *14*, 523-528.
- [14] aM. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *Journal of Computational Chemistry* **1993**, *14*, 1347-1363; bM. S. Gordon, M. W. Schmidt, in *Theory and applications of computational chemistry* (Eds.: C. E. Dykstra, G. Frenking, K. S. Kim, G. E. Scuseria), Elsevier, **2005**.
- [15] V. Barone, M. Cossi, J. Tomasi, *The Journal of Chemical Physics* **1997**, *107*, 3210-3221.
- [16] P. Su, H. Li, *The Journal of Chemical Physics* **2009**, *130*, 074109.
- [17] aM. M. Francel, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, J. A. Pople, *The Journal of Chemical Physics* **1982**, *77*, 3654-3665; bV. A. Rassolov, J. A. Pople, M. A. Ratner, T. L. Windus, *The Journal of Chemical Physics* **1998**, *109*, 1223-1229.
- [18] N. B. Balabanov, K. A. Peterson, *The Journal of Chemical Physics* **2005**, *123*, 064107.
- [19] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *The Journal of Chemical Physics* **1980**, *72*, 650-654.
- [20] A. Albert, *Physical Methods in Heterocyclic Chemistry*, **1963**.
- [21] X. Wang, S. M. Berry, Y. Xia, Y. Lu, *J. Am. Chem. Soc.* **1999**, *121*, 7449-7450.
- [22] C. R. Andrew, R. Fraczekiewicz, R. S. Czernuszewicz, P. Lappalainen, M. Saraste, J. Sanders-Loehr, *Journal of the American Chemical Society* **1996**, *118*, 10436-10445.
- [23] S. Bhattacharya, J. T. Lecomte, **1997**, *73*, 3241-3256.
- [24] E. Arbely, T. J. Rutherford, T. D. Sharpe, N. Ferguson, A. R. Fersht, *Journal of Molecular Biology* **2009**, *387*, 986-992.

Table 3.1 Homo-CPCM/B3LYP/6-31G* optimized geometric parameters (Å and degree)
for the Cu_A model molecules.

	1CC3 chain A	EXAFS ^[9]		model molecule			
		Low pH	High pH	141	142	144	145
O _W -Cu _O	5.11					4.93	2.12
O _W -H ¹ (His120)							1.93
H _W -O(His120)						1.93	1.70
H _W -O(Cys112)						1.91	1.66
Cu _O -Cu _S	2.42	2.36	2.40	2.39	2.29	2.39	2.34
Cu _S -N ^{δ1} (His46)	2.01	1.95	1.92	1.99	1.97	1.98	1.97
Cu _O -N ^{δ1} (His120)	2.06			2.01	3.58	2.00	4.42
Cu _O -S(Cys112)	2.46	2.26	2.27	2.28	2.20	2.29	2.26
Cu _S -S(Cys112)	2.42			2.26	2.27	2.27	2.26
Cu _O -S(Cys116)	2.30			2.27	2.26	2.27	2.27
Cu _S -S(Cys116)	2.29			2.26	2.29	2.26	2.26
Cu _O -O(Glu114)	2.17			2.16	1.93	2.10	2.00
Cu _S -S(Met123)	2.98			2.78	2.64	2.72	2.73
∠Cu _O -Cu _S -N ^{δ1} (His46)	168			148	138	150	149
∠Cu _S -Cu _O -N ^{δ1} (His120)	155			142	125	149	142

Table 3.2 Homo- and Het-CPCM/B3LYP/6-31G* computed pK_a of His120 of Cu_A azurin at 298K.

ϵ	78.39	20	10	4
Homo-CPCM	3.3	1.9	0.0	-5.4
Het-CPCM	3.3	2.5	1.4	-1.7

Table 3.3 Solvation energies (kcal/mol) and relative energies between base-form and acid-form model molecules for 1CC3.

	ϵ	Solvation energies		Relative energies
		144-atom	145-atom	
Homo-CPCM/B3LYP/6-31G*	78.39	-96.9	-141.0	285.91
	20	-92.3	-134.5	283.93
	10	-86.3	-125.9	281.37
	4	-69.0	-101.1	273.90
Het-CPCM/B3LYP/6-31G*	20	-94.8	-137.8	284.78
	10	-92.0	-133.5	283.30
	4	-83.9	-121.2	279.06

Figure 3.1 144-atom model molecule extracted from X-ray structure 1CC3.

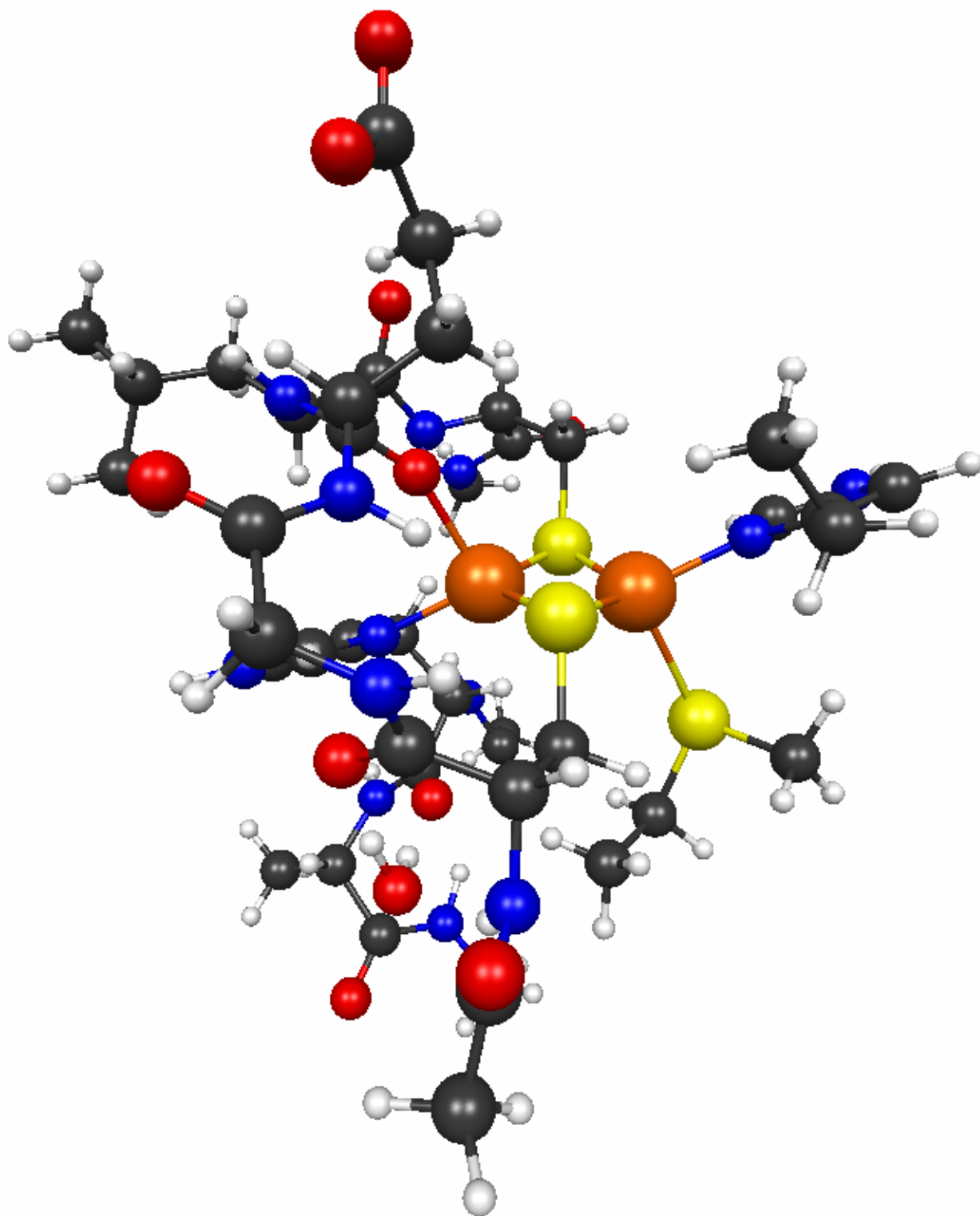
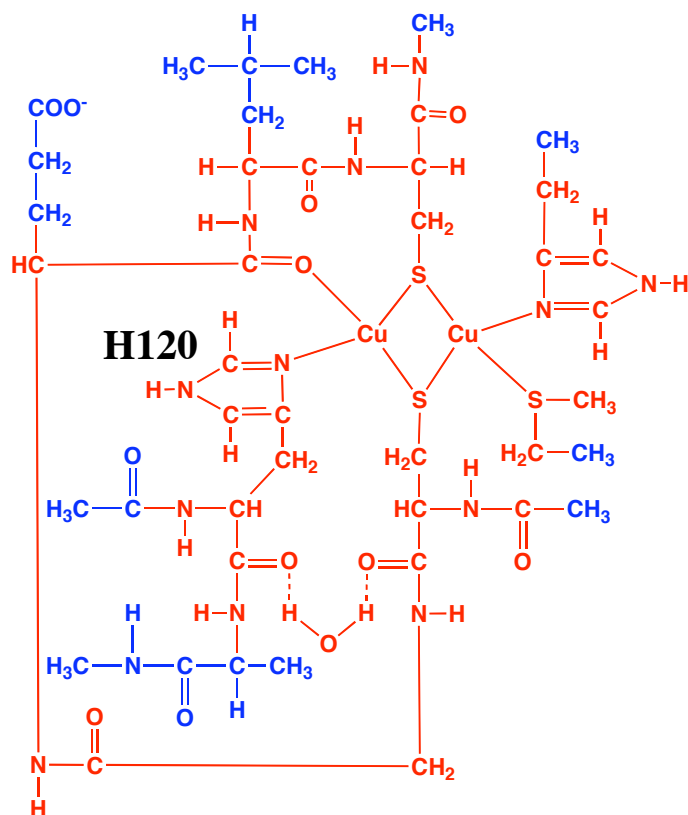
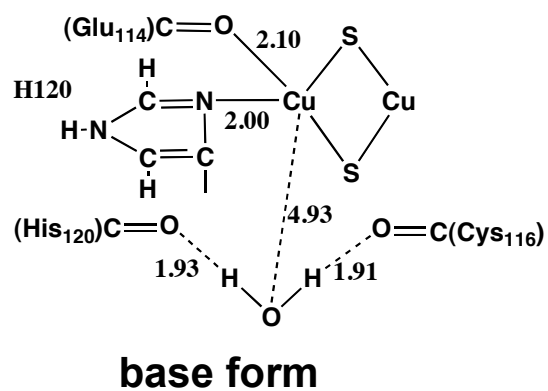


Figure 3.2 Schematic structure of the 144-atom model molecule of the Cu_A center

extracted from X-ray structures 1CC3. 58 atoms (blue) are fixed when 86 atoms (red) are optimized with CPCM/B3LYP/6-31G* method.





Chapter 4

Computational Accuracy and Efficiency in NMR Chemical Shielding

Abstract

We investigated the accuracy of ^{13}C NMR chemical shielding calculations over a database of 19 ^{13}CO -referenced experimental low-pressure ^{13}C gas-phase chemical shieldings, spanning the full carbon chemical shielding range. We compared Hartree Fock (HF), density functional theory (DFT), Møller-Plesset second-order perturbation theory (MP2) and coupled cluster singles doubles, and perturbative triples (CCSD(T)) using Dunning's convergent basis set series aug-cc-pV n Z ($n = 2 - 5$ and $n = 3$ for CCSD(T)) ; and basis set extrapolation using the aug-cc-pV n Z and aug-cc-pCV n Z ($n = 2 - 5$) basis sets, with vibrational corrections at the MP2 level, and electron correlation using coupled cluster methods. After including electron correlation, complete basis set extrapolation, and vibrational correction, the best set of computed results with respect to experimental chemical shielding has a root mean square deviation (RMSD) from experiment of 1.6 ppm, with a first order constant deviation of 0.49 ppm and a linear regression coefficient of 1.007. Basis set incompleteness is found to be the main source of inaccuracy and cannot be removed by applying any fixed correction, but is dependent on the chemical nature of the relevant group. The ^{13}C chemical shielding of methyl, ethylene and ethyne carbons is significantly improved by vibrational correction.

4.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a primary tool for investigating the structure and dynamics of chemical and biochemical molecules. Chemical shielding is the primary NMR parameter in NMR spectroscopy since it has a close relationship with molecular electronic structure and the chemical environment of the molecule. Chemical shielding is an excellent probe for distinguishing different electronic environment inside molecules.

The standard theoretical analysis of the chemical shielding breaks it down into two terms. The diamagnetic term accounts for shielding of the field at the nucleus by induced currents in the orbital electrons. The paramagnetic term derives from mixing into the zero-field ground state tiny quantities of excited electronic configurations with non-zero orbital angular momentum, as a result of the magnetic field.

NMR shielding constants for different nuclei are widely studied by different quantum mechanical methods since the gauge problem^[1], which bedeviled computed magnetic properties because it introduced a strong dependence on axis position and slow convergence, was solved by the introduction and implementation of gauge-including atomic orbital methods (GIAO), which made feasible the use of nucleus centered origins for atomic orbitals.^[2] Because ^{13}C is spin $\frac{1}{2}$ and is present in biological organisms and many other interesting materials, its chemical shielding has been widely used as a probe of structure in many different fields of chemistry.

Three major factors need to be considered to calculate shielding constants in good agreement with experimental values. Because measured NMR chemical shielding is a statistical average of the chemical shielding of all accessible vibrational states of the

molecule, an ensemble average of the shielding over these states needs to be performed. The adequacy of corrections for electron correlation and basis set truncation also has to be assessed. All of these three major problems need to be addressed in any highly accurate NMR chemical shielding calculation.

Gauss and co-workers tested the above factors for small molecules ^{13}C chemical shielding calculations.^[3] Ochsenfeld tested more molecules and narrows the testing factors.^[4]

In the current work, a database of CO-referenced experimental low-pressure ^{13}C gas-phase chemical shifts from 15 small molecules was used to evaluate accuracy of these methods. The database includes CH_3F , CH_4 , CF_4 , HCN , CH_3CN , CH_3NH_2 , CH_2CH_2 , CH_3CH_3 , CH_2CCH_2 , HCCH , CH_3OH , CH_3CHO , CO_2 , and $(\text{CH}_3)_2\text{CO}$. The experimental ^{13}C chemical shielding of the molecules are obtained from the work of A. Keith Jameson and Cynthia J. Jameson^[5]; except for formaldehyde, methane and acetonitrile, which were measured by ourselves. Two families of basis sets, the Dunning augmented correlation-consistent polarized valence n -tuple zeta series^[6] (aug-cc-pV n Z, $n = 2 - 5$) and the core-valence compensated version of those sets^[7] (aug-cc-pCV n Z, $n = 2 - 5$), were selected since the correlation consistent basis sets were initially developed by using methods that recover the correlation energy. Various quantum mechanical methods, from Hartree Fock (HF), to Møller-Plesset second order perturbation theory (MP2),^[8] to coupled-cluster singles, doubles and perturbative-triples (CCSD(T))^[9] methods, as well as density functional theory (DFT) using the hybrid functional of Becke and Lee, Yang and Parr (B3LYP)^[10], which is the most popular functional of DFT calculations of chemical shielding, were chosen, although some other functionals such as KT3,^[11]

OPBE^[12] was recognized to have better performance among DFT methods. We have also corrected the shielding constants by a variational approach to vibrational motion using the potential energy surface and shielding surface by fitting quantum mechanical calculated molecular energies and shielding at different geometries, which is more accurate than previously widely used variation-perturbation approach.^[13]

4.2 Computational Details

The molecules were first fully optimized by using MP2/aug-cc-pVTZ under the GAMESS software package. The optimized equilibrium geometry was used to perform normal mode analysis, then vibrational correction, complete basis set correction and electron correlation correction.

Complete basis set (CBS) chemical shielding (σ_∞) at the equilibrium geometry was obtained by computing chemical shielding using GIAO method at the MP2/aug-cc-pVnZ ($n = 2 - 5$) in the GAUSSIAN 09 software package since correlation consistent basis sets generate a sequence of basis sets that converges toward the basis set limit systematically. Several different schemes have been proposed to compute the infinite basis set limit chemical shielding.^[14] In current research, the computed chemical shielding is fit to the empirical equation:

$$\sigma = \sigma_\infty + Be^{-Cn}$$

where the B parameter is metric of the magnitude of the basis set incompleteness effect and the C parameter of the rapidity of convergence.

Computing vibrational corrections will follow the procedure: The molecules at equilibrium geometry were displaced along each computed normal mode between

displacements $-1/\alpha$ and $1/\alpha$ with a total of up to 21 points. The potential energy surface and shielding surface of each individual vibrational mode j obtained using MP2/aug-cc-pVTZ under the GAUSSIAN 09^[15] software package can be fitted to polynomial functions by using a custom-developed Mathematica notebook, which is included as supplementary material. The shielding and potential functions were fitted as

$$\sigma_j(x) = \sigma_{j,\text{exp}} + \sigma_{j,1}x + \sigma_{j,2}x^2 + \sigma_{j,3}x^3 + \sigma_{j,4}x^4 + \sigma_{j,5}x^5 + \sigma_{j,6}x^6$$

$$V_j(x) = V_{j,\text{exp}} + V_{j,1}x + V_{j,2}x^2 + V_{j,3}x^3 + V_{j,4}x^4 + V_{j,5}x^5 + V_{j,6}x^6$$

where σ_{exp} and V_{exp} are chemical shielding and energy at variationally determined equilibrium points and x is the deviation of normal coordinates j from the equilibrium point, which is determined by the zero of the first derivative of the potential energy function, that is, the minimum energy.^[16]

Linear variation theory is used to figure out chemical shielding vibrational correction.

The harmonic oscillator wave functions ψ_{n_i} were used as the basis set.

$$\psi_{n_i} = N_{n_i} H_{n_i} \left(\frac{x_i}{\alpha_i} \right) e^{-\frac{1}{2} \left(\frac{x_i}{\alpha_i} \right)^2}$$

where $\alpha_i = \left(\frac{\hbar^2}{\mu_i k_i} \right)^{\frac{1}{4}}$ with k_i as force constant and μ_i as reduced mass for mode i . All the parameters are from Hessian calculations. N_{n_i} is a normalization constant and H_{n_i} is a Hermite polynomial.

The trial function is taken to

$$\Psi = \sum_{n_i=1} c_i \psi_{n_i}$$

The vibrational full Hamiltonian is used as

$$\mathbf{H} = -\frac{\hbar^2}{2\mu} \frac{d^2}{dx^2} + V$$

The vibrational eigenvalues, which can be used to compare vibrational frequencies from Hessian calculation to test the accuracy of the calculation, and wavefunctions, which are the eigenfunctions, are determined by solving the Schrödinger equation. Finally, the vibrational correction to chemical shielding can be expressed as

$$\sigma_0 = \sum_{j=1}^N \left(\langle \Psi | \sigma_j | \Psi \rangle - \sigma_{j,\text{exp}} \right)$$

where N is the number of normal modes.

More accurate electron correlation effects were obtained by comparing CCSD(T) and MP2 calculated chemical shielding by using aug-cc-pVTZ using the CFOUR software, which was developed to allow high-level coupled cluster calculations.^[17]

$$\sigma_{EC} = \sigma_{CCSD(T)} - \sigma_{MP2}$$

Our best estimate of the chemical shielding considers electron correlation effects, ``basis set limit effect, and vibrational corrections, and can be expressed as following:

$$\sigma = \sigma_{\infty} + \sigma_0 + \sigma_{EC}$$

The experimental chemical shielding was obtained by using ^{13}C chemical shielding of CO as a reference. The chemical shielding of formaldehyde, acetonitrile and methane was measured by us and will be published separately. The experimental chemical shielding of the other molecules are from the work of A. Keith Jameson and Cynthia J. Jameson, and was adjusted by 0.6 ppm since our own calculations for carbon monoxide, employing a complete basis set extrapolation from computations at the CCSD(T)/aug-cc-pVnZ level, using a second-order ro-vibrational correction to the vibrational ground state over a

thermal ensemble of rotational states, yields a ^{13}C shielding of 1.6 ppm, which is slightly different from reported 1.0 ± 1.2 ppm.^[5]

4.3 Formaldehyde gas phase ^{13}C NMR experiments

Formaldehyde gas was prepared by heating ^{13}C labeled paraformaldehyde at 120 °C overnight in an evacuated J-Y tube^[18]. The tube was then pressurized up to 2 bar with unlabeled CO gas, and ^{13}C NMR spectra collected at 9.4 T. Both ^1H coupled and decoupled ^{13}C spectra, are shown in Figure 1; the CO resonance is clearly seen at the right of the triplet or singlet. The chemical shielding of CO was set at its fully rovibrationally corrected CCSD(T)/aug-cc-pV6Z value of 1.6 ppm. Methane and acetonitrile samples were obtained commercially.

4.4 Results and Discussions

MP2 at the aug-cc-pVTZ level is generally recognized to produce accurate geometries, and so we do not further consider the accuracy of optimizations^[19] Optimized geometries from this method were used as a baseline to perform NMR chemical shielding calculations since it is more accurate than DFT and less expensive than coupled cluster methods.

Electron correlation

Table 1-3 lists the computed chemical shielding for all the 15 molecules by using MP2/aug-cc-pVnZ, HF/aug-cc-pVnZ, and B3LYP/aug-cc-pVnZ, for comparison, the experimental chemical shielding is listed also. All three tables show that the computed chemical shielding decreases with increasing basis set size at MP2, HF and B3LYP

levels. The computed chemical shielding by using MP2 is most accurate; and the best results are obtained using CBS extrapolation in Table 1. The deviation of ^{13}C chemical shielding of methyl groups from experimental value by using MP2/CBS is about 4-6 ppm. The ^{13}C chemical shielding of CH_4 , C_2H_2 and C_2H_4 and allene also have big deviations. However, the deviation is smaller for carbons with high electronegativity atoms such as N, O, F attached. Table 2 and 3 shows that the agreement of computed chemical shielding and experimental shielding is better at lower basis set than at higher basis set when HF and B3LYP was used; this of course is merely a cancellation of errors. Table 4 and 5, respectively, shows RMSD and regression constants for different combination of quantum mechanical methods and basis sets. MP2/CBS gives best RMSD and regression constant in these computations. B3LYP/aug-cc-pVDZ and HF/aug-cc-pVTZ, respectively, have best performance with a RMSD of 4.5 ppm and constant offset of -1.7 ppm for B3LY; and RMSD of 7.7 ppm and constant offset of 6.4 ppm for HF, however the linear regression coefficient of B3LYP/aug-cc-pVDZ is only 0.8. Higher basis sets make computational result much worse by using B3LYP and HF. The above fact may explain why B3LYP is popular currently: since it is generally employed at low basis set levels for routine NMR chemical shift calculations, because it is cheaper and the only available method for big molecules and for lattice calculations, but its success seems to be principally due to partial cancellation of errors. DFT methods in general do not provide a path for systematic improvement in the accuracy of calculation of electron correlation; but in NMR calculations, DFT is even worse; it frustrates any improvement that could be obtained by expanding the basis set.

Coupled cluster CCSD(T) and CCSDT was found to give best results.^[20] Actually CCSD(T) is even better than CCSDT, however, we can not do CBS calculations for most of the molecules by using such big basis sets due to computational resource limits. Fortunately, we find that at moderate basis set levels, CCSD(T) corrections are additive to MP2, independent of the basis set. This can be established using the smallest molecules, for which CCSD(T) basis set extrapolation is possible. We have therefore computed chemical shielding by using coupled cluster methods at the CCSD(T)/aug-cc-pVTZ level to obtain more accurate electron correlation effects. Table 8 shows that the calculated differences between MP2 and CCSD(T) at the aug-cc-pVTZ level are within 2 ppm, except for highly correlated systems such as CO₂ and center carbon of allene, which is about 4 ppm. The following discussions only consider MP2 calculations.

Basis set convergence

Figure 2 shows the CBS extrapolations of four molecules. While the overall shape of the functions is the same, their vertical extents (parameter B) vary widely; the rate of convergence as a function of basis set level also varies, but less obviously. These graphs indicate that a fixed adjustment to the computed chemical shielding will only very poorly correct for the basis set incompleteness problem. The difference between aug-cc-pVDZ and CBS is only 9.5 ppm for methane, but as large as 18.7 ppm for hydrogen cyanide, which has approximately the same B parameter. The difference is even larger with about 28 ppm for the carbonyl carbon of CH₂O, CH₃CHO and (CH₃)₂CO (Table 1). It is clear the computed chemical shielding is not fully converged at $n = 3$ and the difference

between $n = 3$ and CBS is not constant, which is higher for more strongly electron correlated molecules.

Plotting the ^{13}C B and C parameters obtained from CBS extrapolations for 18 carbons in 15 molecules that could be computed at the MP2/aug-cc-pV n Z level with n up to 5 (except acetone), (sufficient to give a reasonable extrapolation) yields the scatter plot showing in Figure 3. The B parameter varies over a wide range, from 90 ppm in CH_4 to 540 ppm for the center carbon in allene. There is an approximate correlation with the nominal hybridization at the carbon; sp^3 carbons have values in the range 90-180 ppm, sp^2 carbons from 200-280 ppm, sp from 250 – 350 ppm and the sole cumulene ($\text{C}=\text{C}=\text{C}$) carbon has a B parameter of 530 ppm. There is also dispersion of the C parameter, but it is much less correlated with hybridization. Both the magnitude of the truncation error B and the rate of convergence of the CBS extrapolation C parameters are highly variable, but strongly clustered for chemically similar carbons in the same functional group, although the data are really too sparse to say so definitively. Aliphatic residues tend to have rather small truncation errors but converge rather irregularly. sp^2 hybridized carbons converge slowly; sp and allene more rapidly. Basis set truncation is serious for unsaturated carbons and most severe for the central carbon of allene.

The basis set aug-cc-pCV n Z was also examined (Table 6 and 7). The chemical shielding differences between CBS of aug-cc-pCV n Z and aug-cc-pV n Z are within 0.5 ppm, which indicate that aug-cc-pV n Z basis set is excellent for the calculation purpose, although aug-cc-pCV n Z converges faster than aug-cc-pV n Z for all the tested molecules, probably simply because of the increased basis set size. The differences between CBS and aug-cc-

pV5Z are generally within 1 ppm, leading us to recommend that the aug-cc-pV5Z basis set is equivalent to complete basis set extrapolation.

Vibrational correction

The vibrational chemical shielding curve is complicated due to the different effects of the vibrational modes on the electronic and bonding structures. The chemical shielding behavior is determined by the symmetry and position of the nucleus, and the type of normal modes. The chemical shielding curve as a function of displacement is close to linear for fully symmetric vibrational modes; it is quadratic for non-fully one dimensional vibrational modes and three dimensional vibrational modes. It is symmetric for two dimensional vibrational modes, except for the C_3 point group, since two dimensional E modes do not match with three dimensional symmetry.

Vibrational corrections are obtained by summing the corrections from each individual vibrational mode. 20 data points are required for fully symmetric modes and 10 data points for non-fully-symmetric modes, because they contain only even terms. The vibrational corrections from degenerate E modes are identical. Although different vibrational modes sometimes yield corrections of opposite sign,^[21] total vibrational corrections to the ^{13}C chemical shielding are uniformly negative (Table 9), which is consistent with literature.^[3] Vibrational corrections of methyl group in CH_3F , CH_3OH , $(\text{CH}_3)_2\text{CO}$, CH_3NH_2 , CH_3CH_3 , CH_3CHO and CH_3CN are biggest, in the range -2.0 ppm ~ -3.6 ppm, while carbons with stronger electron correlation paradoxically have smaller vibrational corrections. The magnitude of the chemical shielding vibrational effects has

no relationship with the potential anharmonicity and the harmonic frequency, which is related with the vibrational correction on geometry.

Overall performance and conclusion

Table 8 shows the overall theoretical result in comparison with the experimental data.

Inclusion of vibrational correction reduces the RMSD of the full data set (computed minus experimental for 19 residues) from 3.3 ppm, obtained using complete basis set extrapolation, to 1.9 ppm; expansion of the electron correlation to the coupled-cluster level further reduced the RMSD to 1.6 ppm. At this level, the absolute agreement with experiment is also best; with a constant data offset of 0.49 ppm and a linear regression coefficient of 1.007, as shown in Table 8 and Figure 4. The only real outlier is the central carbon in allene, for which electron correlation effects are notoriously large. We also find that basis set incompleteness is the dominant limitation for highly accurate chemical shielding calculations; correcting for this, our MP2 computations are comparable with Gauss's work at the coupled cluster level.^[3]

References

- [1] F. London, *J. Phys. Radium* **1937**, 8, 397-409.
- [2] aK. Wolinski, J. F. Hinton, P. Pulay, *Journal of the American Chemical Society* **1990**, 112, 8251-8260; bR. Ditchfield, *Molecular Physics* **1974**, 27, 789-807.
- [3] A. A. Auer, J. Gauss, J. F. Stanton, *Journal of Chemical Physics* **2003**, 118, 10407-10417.
- [4] D. Flaig, M. Maurer, M. Hanni, K. Braunger, L. Kick, M. Thubauville, C. Ochsenfeld, *Journal of Chemical Theory and Computation* **2014**, 10, 572-578.
- [5] A. K. Jameson, C. J. Jameson, *Chemical Physics Letters* **1987**, 134, 461-466.
- [6] T. H. Dunning, *The Journal of Chemical Physics* **1989**, 90, 1007-1023.
- [7] D. E. Woon, T. H. Dunning, *The Journal of Chemical Physics* **1995**, 103, 4572-4585.
- [8] aJ. Gauss, *The Journal of Chemical Physics* **1993**, 99, 3629-3643; bM. Kollwitz, M. Häser, J. Gauss, *The Journal of Chemical Physics* **1998**, 108, 8295-8301.
- [9] J. Gauss, J. F. Stanton, *Chemical Physics Letters* **1997**, 276, 70-77.
- [10] aA. D. Becke, *The Journal of Chemical Physics* **1993**, 98, 5648-5652; bC. Lee, W. Yang, R. G. Parr, *Physical Review B* **1988**, 37, 785-789.
- [11] D. O. Samultsev, V. A. Semenov, L. B. Krivdin, *Magnetic Resonance in Chemistry* **2014**, 52, 222-230.
- [12] Y. Zhang, A. N. Wu, X. Xu, Y. J. Yan, *Chemical Physics Letters* **2006**, 421, 383-388.
- [13] aP.-O. Åstrand, K. Ruud, P. R. Taylor, *The Journal of Chemical Physics* **2000**, 112, 2655-2667; bP.-O. Åstrand, K. Ruud, D. Sundholm, *Theor Chem Acc* **2000**, 103, 365-373.
- [14] aP. Norman, Y. Luo, D. Jonsson, H. Ågren, *The Journal of Chemical Physics* **1997**, 106, 8788-8791; bA. Karton, J. L. Martin, *Theor Chem Acc* **2006**, 115, 330-333.
- [15] aM. J. Frisch, G. W. Trucks, H. B. Schlegel, e. al., *GAUSSIAN 09, Revision A.02* **2009**; bM. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *Journal of Computational Chemistry* **1993**, 14, 1347-1363.
- [16] P. O. Åstrand, G. Karlström, A. Engdahl, B. Nelander, *The Journal of Chemical Physics* **1995**, 102, 3534-3554.
- [17] J.F. Stanton, J. Gauss, M.E. Harding, et.al., *CFOUR, a quantum chemical program package*. For the current version, see <http://www.cfour.de>.
- [18] D.D.P, W.L.F.A, D.R.P, in *Purification of Laboratory Chemicals (Sixth Edition)* (Ed.: W. L. F. A. L. L. Chai), Butterworth-Heinemann, Oxford, **2009**, p. iv.
- [19] K. B. Wiberg, *Journal of Computational Chemistry* **2004**, 25, 1342-1346.
- [20] aJ. Gauss, *Journal of Chemical Physics* **2002**, 116, 4773-4776; bJ. Gauss, J. F. Stanton, in *Advances in Chemical Physics, Vol 123, Vol. 123*, **2002**, pp. 355-422.
- [21] H. Sabzyan, B. Buzari, *Chemical Physics* **2008**, 352, 297-305.

Figure 4.1 Gas phase formaldehyde of ^1H coupled and decoupled ^{13}C NMR spectra.

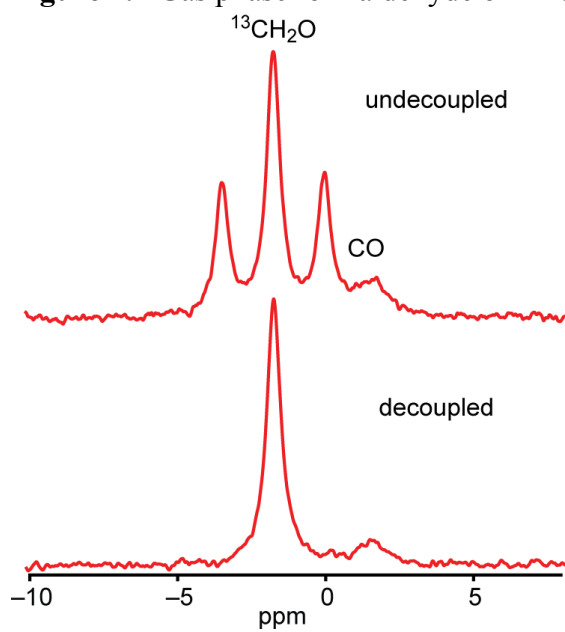


Table 4.1 ^{13}C NMR chemical shielding of molecules, in ppm, at the MP2 level

Compound	aug-cc-pVDZ	aug-cc-pVTZ	aug-cc-pVQZ	aug-cc-pV5Z	CBS	Experiment ^b
CH_3F	136.4	124.8	121.3	119.9	119.4	117.4
CH_4	209.8	203.4	201.4	200.7	200.4	196 ^c
CH_3OH	155.0	144.7	141.6	140.4	140.0	137.2
$(\text{CH}_3)_2\text{CO}$	176.4	166.9	164.4		163.4	158.6
$(\text{CH}_3)_2\text{CO}$	15.6	-2.8	-8.5		-11.1	-12.5
CF_4	80.3	68.0	64.2	62.3	61.7	65.1
CH_3NH_2	176.6	167.2	164.7	163.6	163.4	158.9
CH_3CH_3	197.5	189.9	187.8	187.0	186.8	181.5
CH_3CHO	174.4	165.2	162.6	161.5	161.2	157.8
CH_3CHO	23.0	4.8	-1.0	-3.9	-5.0	-6.1
CH_3CN	201.1	196.0	193.2	192.8	192.8	188.6 ^c
CH_3CN	92.0	80.1	74.3	73.3	73.6	74.8 ^c
CO_2	79.6	69.6	63.0	62.0	61.2	59.4
HCN	106.8	91.2	87.2	85.2	84.9	82.7
CH_2CCH_2	6.1	-20.3	-25.7	-28.9	-28.8	-28.7
CH_2CCH_2	138.2	124.7	121.2	119.4	119.2	115.8
HCCH	143.1	127.5	124.5	122.7	122.8	117.8
H_2CCH_2	92.3	75.4	71.0	68.7	68.3	65.1
CH_2O	28.5	10.3	4.3	1.3	0.2	-1.0 ^c

- All calculated ^{13}C NMR chemical shielding using ppm as unit.
- Experimental data are from Jameson's work
- Experimental data are from our own research.

Table 4.2 ^{13}C NMR chemical shielding of molecules, in ppm, at the B3LYP level.

Compound	aug-cc-pVDZ	aug-cc-pVTZ	aug-cc-pVQZ	aug-cc-pV5Z	CBS	Experiment
CH_3F	122.8	110.7	106.2	104.5	103.5	117.4
CH_4	198.8	191.9	189.2	188.3	187.6	196
CH_3OH	141.4	130.5	126.5	124.9	124.0	137.2
$(\text{CH}_3)_2\text{CO}$	163.3	153.5	150.2	148.8	148.3	158.6
$(\text{CH}_3)_2\text{CO}$	-14.4	-30.8	-38.1	-41.4	-44.1	-12.5
CF_4	63.6	51.5	46.1	43.8	42.0	65.1
CH_3NH_2	162.9	152.7	149.3	147.9	147.3	158.9
CH_3CH_3	184.2	175.7	172.8	171.7	171.2	181.5
CH_3CHO	160.5	150.8	147.5	146.1	145.4	157.8
CH_3CHO	-7.3	-23.3	-30.3	-33.6	-36.2	-6.1
CH_3CN	190.8	183.1	180.4	179.4	178.9	188.6
CH_3CN	78.1	62.2	57.0	54.5	53.5	74.8
CO_2	63.4	53.2	46.8	44.8	41.4	59.4
HCN	85.8	71.9	67.1	64.8	63.7	82.7
CH_2CCH_2	-20.1	-45.1	-52.4	-56.1	-57.1	-28.7
CH_2CCH_2	122.1	109.1	104.7	102.7	101.9	115.8
HCCH	127.1	111.1	107.3	105.2	105.0	117.8
H_2CCH_2	69.8	53.9	48.4	45.8	44.6	65.1
CH_2O	-3.7	-19.2	-26.1	-29.3	-31.9	-1.0

Table 4.3 ^{13}C NMR chemical shielding of molecules, in ppm, at the HF level.

Compound	aug-cc-pVDZ	aug-cc-pVTZ	aug-cc-pVQZ	aug-cc-pV5Z	CBS	Experiment
CH_3F	137.6	127.1	124.3	123.3	123.0	117.4
CH_4	204.1	197.2	195.3	194.7	194.5	196
CH_3OH	155.4	145.6	143.1	142.1	141.9	137.2
$(\text{CH}_3)_2\text{CO}$	173.9	165.4	163.4	162.6	162.4	158.6
$(\text{CH}_3)_2\text{CO}$	-8.2	-20.7	-25.0	-27.3	-28.2	-12.5
CF_4	91.7	81.8	78.8	77.4	76.9	65.1
CH_3NH_2	174.5	165.4	163.2	162.3	162.2	158.9
CH_3CH_3	192.5	185.2	183.4	182.7	182.6	181.5
CH_3CHO	172.3	164.0	161.9	161.1	160.9	157.8
CH_3CHO	0.2	-12.1	-16.3	-18.5	-19.4	-6.1
CH_3CN	199.9	192.6	190.9	190.3	190.2	188.6
CH_3CN	77.5	64.1	60.8	59.1	58.8	74.8
CO_2	62.7	52.7	48.4	47.0	45.8	59.4
HCN	84.5	71.5	68.6	66.9	66.8	82.7
CH_2CCH_2	-20.8	-40.0	-44.2	-46.8	-46.9	-28.7
CH_2CCH_2	128.2	117.4	114.8	113.4	113.2	115.8
HCCH	130.7	117.6	115.4	113.9	144.1	117.8
H_2CCH_2	75.9	63.3	60.0	58.2	57.9	65.1
CH_2O	7.1	-5.2	-9.3	-11.4	-12.3	-1.0

Table 4.4 ^{13}C NMR chemical shielding RMSD with respect to experimental value.

Methods	aug-cc-pVDZ	aug-cc-pVTZ	aug-cc-pVQZ	aug-cc-pV5Z	CBS
HF	12.8	7.7	8.8	9.7	11.6
MP2	22.1	8.4	4.8	3.5	3.3
B3LYP	4.5	10.8	15.8	18.0	19.4
CCSD(T)		8.3			

Table 4.5 ^{13}C NMR chemical shielding offset from linear regression fit with respect to experimental value.

Methods	aug-cc-pVDZ	aug-cc-pVTZ	aug-cc-pVQZ	aug-cc-pV5Z	CBS
HF	-6.9	6.4	10.1	12.1	12.3
MP2	-30.3	-9.1	-3.4	-0.6	0.35
B3LYP	-1.7	14.9	20.9	23.7	25.4
CCSD(T)		-9.7			

Table 4.6 ^{13}C NMR chemical shielding of molecules at MP2 level with ppm as unit.

Compound	aug-cc-pCVDZ	aug-cc-pCVTZ	aug-cc-pCVQZ	aug-cc-pCV5Z	CBS	Experiment
CH_3F						117.4
CH_4	207.5	202.1	200.9	200.5	200.5	196
CH_3OH	151.8	142.68	140.7	140.1	140.0	137.2
$(\text{CH}_3)_2\text{CO}$						158.6
$(\text{CH}_3)_2\text{C}\underline{\text{O}}$						-12.5
CF_4	77.9	65.7				65.1
CH_3NH_2						158.9
CH_3CH_3						181.5
CH_3CHO						157.8
$\text{CH}_3\text{C}\underline{\text{H}}\text{O}$						-6.1
CH_3CN	201.1	194.51	193.2	192.8	192.7	188.6
$\text{CH}_3\text{C}\underline{\text{N}}$	92.0	77.5	74.3	73.3	73.2	74.8
CO_2	79.6	66.6	63.0	62.0	61.7	59.4
HCN	102.4	88.9	85.7	84.8	84.6	82.7
$\text{CH}_2\text{C}\underline{\text{H}}\text{CH}_2$	-1.2	-24.1	-28.4		-29.4	-28.7
$\text{CH}_2\text{C}\underline{\text{H}}\text{C}\underline{\text{H}}\text{CH}_2$	134.3	122.3	119.9		119.2	115.8
HCCH	138.5	125.5	123.1	122.4	122.4	117.8
H_2CCH_2	87.2	72.2	69.1		68.2	65.1
CH_2O	23.8	6.3	1.9	0.5	0.1	-1.0

Table 4.7 Vibrational corrections and CCSD(T) electron correlation corrections of ^{13}C NMR chemical shielding of molecules at the MP2/aug-cc-pVTZ level.

Compound	σ_{EC}	σ_0
CH_3F	1.5	-2.0
CH_4	-1.9	-0.8
CH_3OH	1	-3.7
$(\underline{\text{CH}}_3)_2\text{CO}$	-0.3	-3.2
$(\text{CH}_3)_2\underline{\text{C}}\text{O}$	-0.9	-1.3
CF_4	2.2	-0.4
CH_3NH_2	-0.2	-3.7
CH_3CH_3	-1.3	-2.4
$\underline{\text{CH}}_3\text{CHO}$	-0.2	-3.6
$\text{CH}_3\underline{\text{C}}\text{HO}$	-0.9	-1.6
$\underline{\text{CH}}_3\text{CN}$	-1.7	-2.6
$\text{CH}_3\underline{\text{C}}\text{N}$	0.5	-1.6
CO_2	-3.6	-0.5
HCN	-2	-0.4
$\text{CH}_2\underline{\text{C}}\text{CH}_2$	4	-0.7
$\underline{\text{CH}}_2\text{C}\underline{\text{CH}}_2$	-0.7	-1.7
HCCH	-0.5	-2.5
H_2CCH_2	1.3	-2.9
CH_2O	-0.6	-2.5

Table 4.8 Computed ^{13}C NMR chemical shielding of molecules by considering complete basis set effect.

Compound	σ_{CBS}	$\sigma_{\text{CBS}} + \sigma_0$	$\sigma_{\text{CBS}} + \sigma_0 + \sigma_{\text{EC}}$	Experiment
CH_3F	119.4	117.4	118.9	117.4
CH_4	200.4	199.6	197.9	196 ^c
CH_3OH	140	136.3	137.3	137.2
$(\text{CH}_3)_2\text{CO}$	163.4	160.3	159.9	158.6
$(\text{CH}_3)_2\text{CO}$	-11.1	-12.4	-13.2	-12.5
CF_4	61.7	61.3	63.5	65.1
CH_3NH_2	163.4	159.7	159.5	158.9
CH_3CH_3	186.8	184.4	183.2	181.5
CH_3CHO	161.2	157.6	157.5	157.8
CH_3CHO	-5	-6.6	-7.4	-6.1
CH_3CN	192.8	190.2	188.5	188.6
CH_3CN	73.6	72	72.5	74.8
CO_2	61.2	60.7	57.1	59.4
HCN	84.9	84.5	82.5	82.7
CH_2CCH_2	-28.8	-29.5	-25.5	-28.7
CH_2CCH_2	119.2	117.5	116.8	115.8
HCCH	122.8	120.3	119.8	117.8
H_2CCH_2	68.3	65.4	66.7	65.1
CH_2O	0.2	-2.3	-2.9	-1.0
RMSD	3.3	1.9	1.6	
regression constant coefficient	0.35	-0.95	-0.49	
regression linear coefficient	1.022	1.014	1.007	

Figure 4.2 Computed ^{13}C chemical shielding for CH_4 , CH_3OH , H_2CCH_2 and HCN at MP2/aug-cc-pVnZ level, as a function of n .

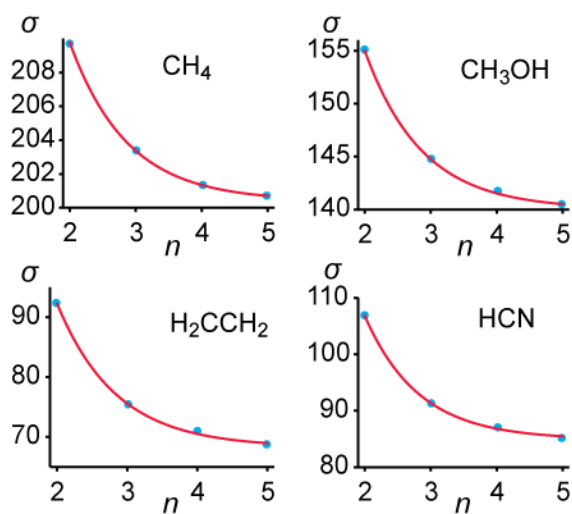


Figure 4.3 Fit parameters B and C for the CBS extrapolation of computed MP2/aug-cc-pVnZ level ^{13}C chemical shieldings for 18 carbons in 15 molecules.

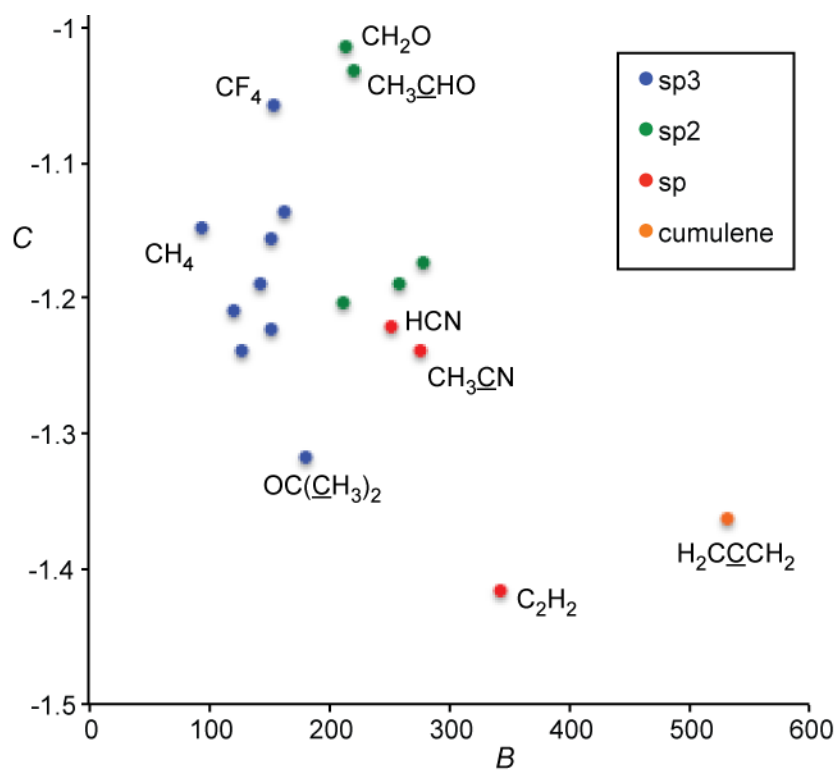
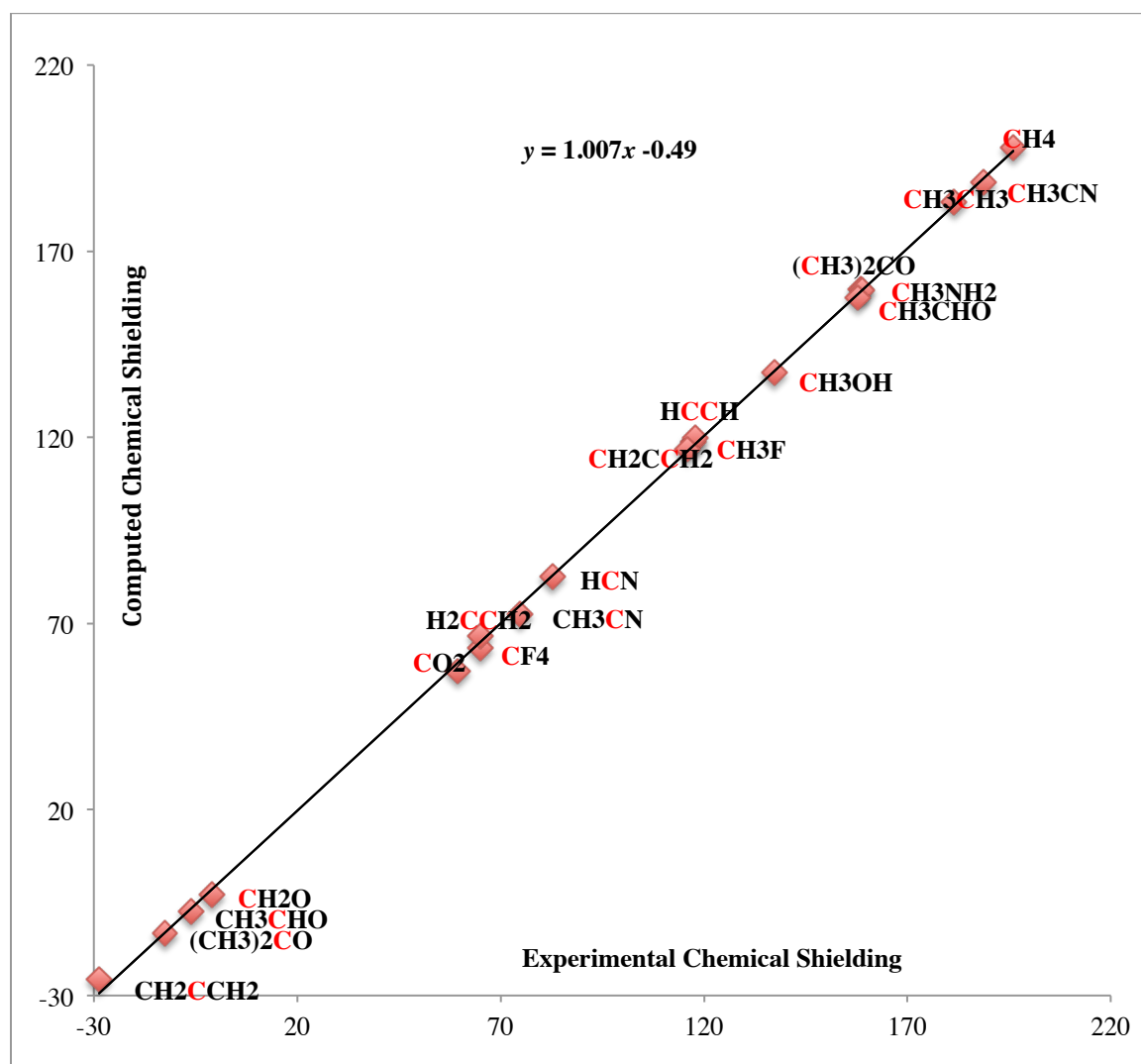


Figure 4.4 The computed chemical shielding, including CBS extrapolation, vibrational correction and electron correlation effects with respect to experimental gas phase chemical shielding for 18 carbons in 15 molecules.



Chapter 5

Thermodynamics and Kinetics of Acid-Catalyzed Retinoic Acid Isomerization

Abstract

The all-*trans*, 13-*cis* and 9-*cis* isomers of retinoic acid all have important endocrine functions in higher organisms, while the 11-*cis* isomer is implicated in vision. The kinetics and thermodynamics of the trifluoroacetic acid catalyzed isomerization of retinoic acid in CD₂Cl₂ were studied by ¹H NMR method. The first three isomers (as well as the 9,13-*dicis* isomer) are all present in substantial amounts in an equilibrium mixture. The all-*trans*, 13-*cis*, 9-*cis* and 9,13-*dicis* all rapidly decay to polymer with rate constants, respectively, to be 22.2, 5.25 and 12.5 d⁻¹, which indicate that retinoic acid decays to polymer when the isomer interconversion happens. Our current study also indicates that all-*trans*, 9-*cis* and 9,13-*dicis* isomers first convert to the 13-*cis* isomer. The above facts indicate that a common intermediate occurs for trifluoroacetic acid catalyzed retinoic acid isomerization. The conversion rate of all-*trans* and 13-*cis* retinoic acid to other isomers are faster than 9-*cis* and 9,13-*dicis*, which is probably due to the lower concentration of 9-*cis* and 9,13-*dicis* relative to all-*trans* and 13-*cis* retinoic acid. The free energy changes of 13-*cis*, 9-*cis* and 9,13-*dicis* relative to all-*trans* retinoic acid 2.11 ± 0.10 , 4.01 ± 0.13 and 7.31 ± 0.7 kJ/mol, respectively. Free energy changes compare well with experimental results.

5.1 Introduction

Retinoids are a class of compounds chemically related to retinol (vitamin A) and oxidation products such as retinals and retinoic acid. Retinoic acid has a large number of geometrical isomers, the most biologically relevant are shown in Figure 1. All contain a trimethylcyclohexene ring attached by a nine-membered conjugated side chain capped by a prosthetic group.

Retinoids fulfill a remarkable variety of roles in living organisms. *All-trans* retinal is the chromophore in the proton pump bacteriorhodopsin and the ion pump halorhodopsins, and its 13-*cis* isomer plays a role in the photocycle and in the dark-adapted form of bacteriorhodopsin. Photoisomerisation of the polyene chain to a 13-*cis* configuration is the first step of a cycle that ultimately results in the translocation of one or two protons across the cell membrane. An 11-*cis* retinal Schiff base is the chromophore of the vertebrate photoreceptors rhodopsin^[1] and iodopsin^[2]. Photoisomerization of the 11-*cis* isomer to the more stable *all-trans* isomer causes detachment of the chromophore and ultimately signals absorption of a photon. More recently, *all-trans* retinoic acid and the 9-*cis* and 13-*cis* isomers play a multiplicity of roles in endocrine function^[3], and in cell differentiation and regulation^[4].

The mechanism of interconversion of these various isomers has long been an active area of research. Early work proposed that the biological generation of 11-*cis* retinal from *all-trans* retinal might be accomplished by chemically or enzymatically catalyzed thermodynamic equilibration^[5]. This now appears thermodynamically implausible, (see

below); A *cis* bond that exchanges a polyene chain for a proton exacts a heavy energetic price. On the other hand, isomers that involve interchanging polyene and methyl substituents about a double bond are more favored. 9-*cis*, 13-*cis* and 9,13-*dicis* isomers have been found to be in equilibrium with all-*trans* in measurable quantities^[6]. Therefore, while the organism can provide itself with adequate quantities of the 9-*cis* and 13-*cis* isomers through a thermodynamic equilibrium, the production of 11-*cis* isomer almost certainly requires energetic input^[7].

Retinoic acid isomerization has been widely studied *in vitro* and *in vivo*. Isomerization of retinoic acid under cell culture conditions was found to result in all-*trans*, 13-*cis*, 9-*cis*, 9,13-*dicis* retinoic acid^[8]. Glutathione S-transferases was found to catalyze retinoic acid Isomerization. The isomerization is stereospecific due to a preference for 13-*cis* retinoic acid from all-*trans* retinoic acid^[9]. All-*trans*, 13-*cis* and 9-*cis* retinoic acid were shown to interconvert in liver microsomes of rats *in vitro*^[10], and 9-*cis* retinoic acid was also observed to be convert to 9,13-*dicis* retinoic acid^[11]. The isomerization of all-*trans* retinoic acid to the 9-*cis* isomer was found to be mediated by thiol radical liver microsomes^[7, 12]. 9-*cis* retinoic acid was produced via isomerization in human mammary epithelial cells^[13]. 13-*cis* retinoic acid was found to rapidly isomerize to all-*trans* retinoic acid on human sebocytes^[14].

Photoisomerization of retinoic acid in various solvents under ambient light^[15], long wavelength UV^[16] and white fluorescent lamps^[17] can be followed by separation of the isomers using high-performance liquid chromatography (HPLC). Retinoic acid can be oxidized and isomerized using iodine and light^[18]. Trifluoroacetic acid (TFA) has been

used as a mobile phase solvent to separate the retinoic acid isomers in HPLC^[7-8, 12, 19] and as catalyst of retinal isomerization^[6].

Previous measurements of the equilibrium concentrations of these isomers by the Rando group have used fast quench HPLC to determine relative amounts of each isomer after isomerization of a solution of a single retinoid species using TFA, iodine^[6], or liver microsomes^[7] as a catalyst. One significant problem with this approach is that it neglects the complicating kinetic effects of side reactions and of the polymerization of retinoic acid. Of particular note, the results of the two studies by the Rando group are not at all consistent.

In the present work, high field 500 MHz ¹H NMR spectroscopy is used to follow the isomerization reactions of all-*trans*, 9-*cis*, 13-*cis* and 9,13-*dicis* retinoic acid, catalyzed by TFA in dichloromethane. The concentration of each isomer was followed without quenching. The approach also accounts for the confounding effect of side reactions. The concentration of each isomer was estimated from NMR resonances within uncluttered regions of the spectrum.

5.2 Materials and Methods

Materials

9-*cis*, 13-*cis*, all-*trans* retinoic acid, CD₂Cl₂, phosphorus pentoxide and trifluoroacetic acid were products of Aldrich Chemical Company. Retinoic acids and CD₂Cl₂ were used without further purification. Dry trifluoroacetic acid was collected at 72.4 °C by distillation of mixture of TFA and phosphorus pentoxide after stirring 6 hours under

nitrogen gas. 5 mm medium wall NMR low vacuum/pressure (LVP) NMR tubes were purchased from Wilmad-Lab Glass.

NMR analysis of isomer structure and isomerization reaction

2D ^1H correlation spectroscopy (COSY) NMR experiment of pure all-*trans*, 13-*cis* and 9-*cis* retinoic acid in a CD_2Cl_2 solution and LVP NMR tube, were performed to assign the vinyl protons of each isomer. The identity of 9,13-*dicis* retinoic acid was established by comparison to a known NMR spectra^[20]. Double quantum filtered correlation spectroscopy (DQF-COSY) ^1H NMR of photoisomerized 13-*cis* retinoic acid was used to assign 9,13-*dicis* retinoic acid vinyl protons. 5.0 mg all-*trans*, 5.0 mg 13-*cis* or 1 mg 9-*cis* retinoic acid in 0.6 mL CD_2Cl_2 were mixed with 0.3 μL dry TFA in a LPV NMR tube and wrapped by aluminum foil to prevent retinoic acid isomerization by light. To study the kinetics and thermodynamics of retinoic acid isomerization, the ^1H NMR spectra of the isomer mixtures, were collected every 15 minutes on a Bruker AVANCE 500 MHz for a total of more than 10 hours.

Data Analysis

Peak positions and intensities were obtained by fitting each multiplet to a sum of Gaussian functions to obtain intensities to extract isomer concentrations.

$$G(x, \omega, d, I) = I e^{-\frac{(x-\omega)^2}{2d^2}} \quad (1)$$

where x is the spectral frequency, ω is the peak frequency, d is the half height width and I is the intensity.

Once intensities were extracted as described above, the entire data set from all three runs was fit to a single set of rate coefficients, making the fit statistically more robust than individual fits would be.

Preliminary inspection indicated the rate equations were all pseudo-first-order. Intensities were therefore fitted to the kinetic scheme in Figure 2. The sixteen rate coefficients were subject to three constraints imposed by the thermodynamics of the system.

$$k_{cb} = \frac{k_{ab} \times k_{bc} \times k_{ca}}{k_{ac} \times k_{ba}} \quad (2a)$$

$$k_{dc} = \frac{k_{ac} \times k_{cd} \times k_{da}}{k_{ad} \times k_{ca}} \quad (2b)$$

$$k_{db} = \frac{k_{ab} \times k_{bd} \times k_{da}}{k_{ad} \times k_{ba}}$$

A matrix for the kinetic scheme can be expressed as:

$$\frac{d\mathbf{A}}{dt} = \mathbf{KA} = \mathbf{K} \begin{pmatrix} [all - trans] \\ [13 - cis] \\ [9 - cis] \\ [9,13 - dicis] \end{pmatrix} \quad (3)$$

where \mathbf{A} is the concentration vector. \mathbf{K} is the rate coefficient matrix and can be expressed as:

$$\mathbf{K} = \begin{pmatrix} -k_{aa} - k_{ab} - k_{ac} - k_{ad} & k_{ba} & k_{ca} & k_{da} \\ k_{ab} & -k_{ba} - k_{bb} - k_{bc} - k_{bd} & k_{cb} & k_{db} \\ k_{ac} & k_{bc} & -k_{ca} - k_{cb} - k_{cc} - k_{cd} & k_{dc} \\ k_{ad} & k_{bd} & k_{cd} & -k_{da} - k_{db} - k_{dc} - k_{dd} \end{pmatrix} \quad (4)$$

\mathbf{K} can be diagonalized. The eigenvector matrix of \mathbf{K} is defined as \mathbf{P} and the eigenvalue matrix as Λ : $\mathbf{P}^{-1} \mathbf{K} \mathbf{P} = \Lambda$. We can now express (3) as

$$\frac{d\mathbf{B}}{dt} = \mathbf{P} \frac{d\mathbf{A}}{dt} = (\mathbf{P} \mathbf{K} \mathbf{P}^{-1}) (\mathbf{P} \mathbf{A}) = \Lambda \mathbf{B} \quad (5)$$

This is now a set of uncoupled linear differential equations

$$\frac{dB_i}{dt} = \Lambda_{ii} B_i \quad (6)$$

with i running from 1 to 4. They have the trivial solution

$$B_i(t) = \exp(\Lambda_{ii} t) B_i(0) \quad (7)$$

and the computed concentrations can be back-calculated using $\mathbf{A}(t) = \mathbf{P}^{-1} \mathbf{B}(t)$. Starting with initial guesses for the values of the thirteen independent rate coefficients, and the four initial concentrations, a least squares minimization was used to compute the coefficients.

$$\chi^2 = \sum_{j=1}^{97} \left[A_{\text{exp},j} - A_{\text{calc}}(t_j | k_1 \dots k_{13} A_1(0) \dots A_3(0)) \right]^2 \quad (8)$$

In above equation, $A_{\text{exp},j}$ are the experimental intensities, and $A_{\text{calc}}(t_j | k_1 \dots k_{13} A_1(0) \dots A_3(0))$ are computed values with $k_1 \dots k_{13}$ being the independent rate coefficients, $A_1(0) \dots A_3(0)$ are

the initial isomer intensities, and t_j are the time at which the intensity was collected. A total of 97 experimental intensities were used. Once the global least squares minimum was achieved, we numerically evaluated the Hessian matrix α composed of the second derivatives of χ^2 , with respect to each pair of parameters. The covariance matrix \mathbf{C} is the inverse of α . This covariance matrix was used to determine standard deviations for each rate coefficient and the equilibrium constant, using standard methods for propagation of errors.

5.3 Results and Discussion

Assignment of ^1H NMR spectra of retinoic acid in CD_2Cl_2

The retinoic acid isomers analyzed in these NMR studies are *9-cis*, *13-cis*, *9,13-dicis* and *all-trans*. The downfield regions of the 1D ^1H spectra of *all-trans*, *13-cis*, and *9-cis* isomers are shown in Figure. 3, 5 and 7. Vinyl ^1H chemical shift assignments of the *all-trans*, *9-cis* and *13-cis* retinoic acid isomers were obtained from the analysis of 2D ^1H COSY spectra (Figures 4, 6, 8) of the pure isomers. ^1H chemical shift assignments of *9,13-dicis* retinoic acid were obtained from the analysis of the DQ-COSY^[21] (Figure 9) of the *13-cis* retinoic acid isomer after photoisomerization. The NMR assignments for the retinoic acid isomers is straightforward: H-14 is always a singlet and H-11 is a doublet of doublets for all retinoic acid isomers; H-10 and H-12 are single- and double-bonded to H-11, and can be distinguished by their J couplings. Weak four-bond couplings from H-12 to H-14 can also be detected in the COSY spectrum. H-7 and H-8 peaks are coupled with each other and the coupling constants are largest, about 16 Hz. The H-7 chemical shift of *all-trans* and *13-cis* is higher than the H-8 chemical shift, In *9-cis* and *9,13-dicis*, H-8

experiences a strong downfield shift and has a higher chemical shift. The chemical shift assignment for H-8 is consistent with a prior NMR assignment for retinoic acid in DMSO using homo- and heteronuclear two-dimensional chemical shift correlation experiments^[22].

Table 1 shows the vinyl proton chemical shift and coupling constants using CD₂Cl₂ (5.33 ppm) as an internal reference. Chemical shifts colored red are equivalent to values previously reported retinoic acid isomers in trifluoroacetic acid. The intensities of these peaks were used to extract isomer concentrations. Only the H-11 and H-12 peaks of 9,13-dicis retinoic acid were assigned due to overlap with other resonances.

Kinetics of retinoic acid isomer interconversion under trifluoroacetic acid

No systematic kinetics data are available for retinoic acid isomerization. In our current study, I₂ was initially used as a catalyst, however isomer decay is too fast and the isomerization can not be detected in detail. Therefore, rate constants were measured by using the trifluoroacetic acid catalyzed interconversion of the all-*trans*, 13-*cis*, 9-*cis* and 9,13-dicis retinoic acid isomers and their decay. Rate coefficients for the complex network of chemical reactions are first order. The rate constants best fit to the NMR data are given in Table 2. All sixteen rate coefficients are significantly different from zero, suggesting that double isomerizations (all-*trans* → 9,13 dicis and 9-*cis* → 13-*cis*) proceed with a single rate-limiting step.

The full ¹H NMR spectrum of all-*trans* retinoic acid is shown in Figure 10. The NMR spectrum in Figure 10 indicates that all-*trans* retinoic acid is completely converted to

sideproducts by TFA after 735 minutes. Therefore, only time points up to 500 min were used in the data analysis.

The change in the NMR peak intensities as function of time for the four isomers converted from all-*trans* retinoic acid are shown in Figure 11. The intensity of all-*trans* retinoic acid rapidly decays with a rate coefficient of 12.5 d^{-1} . By 100 minutes, 13-*cis*, 9,13-*dicis* and 9-*cis* retinoic acid have reached a maximum, and then begin to decay. It is clear that degradation of the all-*trans* isomer is faster than the production of the other isomers; no more than 12% of all-*trans* isomer is converted to the other isomers. It is also noteworthy that there is no time-lag in the generation of 9,13-*dicis* retinoic acid from the all-*trans* isomer, as would be expected if the reaction proceeded through a single bond isomerization. Clearly, nucleophilic addition to odd-numbered carbons, reversing the single/double bond alternation, followed by elimination, is capable of sustaining simultaneous two-bond isomerizations as well as one-bond isomerizations. For example, proton cation from TFA first attach to C-13 of all-*trans* retinoic acid; then follow Michael addition mechanism and produce 9-*cis*, 13-*cis* and 9,13-*dicis* isomers.

TFA catalyzed 13-*cis* and 9-*cis* retinoic acid isomers have similar kinetic behaviors (Figures 12 and 13), but 13-*cis* retinoic acid degrades to side-products faster than the all-*trans* isomer. Conversely, 9-*cis* degrades more slowly than all-*trans*. The reason for this difference is unclear. The fit data (blue curves) agree with experimental points quite well. The 13-*cis*, 9-*cis* isomerization is a synchronous two bond isomerization that shows no time lag and occurs as a single-step process (Figures 14 and 15). The mole fraction of 9,13-*dicis* retinoic acid is never higher than 0.02, which explains why 9,13-*dicis* was

difficult to detect by HPLC^[11]; Rate coefficients for the 9,13-*dicis* isomer have considerably higher errors.

Thermodynamics of trifluoroacetic acid catalyzed retinoic acid isomerization.

The equilibrium constants for 13-*cis*, 9-*cis* and 9,13-*dicis* retinoic acid relative to all-*trans* retinoic acid are 0.430, 0.199 and 0.0523. The standard Gibbs free energies ΔG° are equal to 2.11 ± 0.10 , 4.01 ± 0.13 and 7.31 ± 0.7 kJ/mol, respectively (Table 3). Theoretical ΔG° values, were computed using full thermochemistry under the harmonic approximation, at the B3LYP/6-311++(2d,p) level, and corrected for solvation using a polarizable continuum model.

Our ΔG° values are compared against isomer ratio values for retinal and retinoic acid determined by Rando and Chang^[6-7]. ΔG° value of retinal was calculated by us. ΔG° value of retinoic acid was from literature. Interestingly, the equilibrium percentage of 13-*cis* and 9-*cis* retinoic acid were reported to be different: 17.2 ± 1.1 and 14.8 ± 0.6 , however the ΔG° of 13-*cis* and 9-*cis* retinoic acid were reported to be the same: 3.3 kJ/mol, which cannot be reproduced by using any of the reported percentage at either 25 or 37 Celsius degree. Standard free energies computed from the Rando and Chang ratios for retinals agree better with our numbers than the standard free energies for retinoic acid published by Urbach and Rando.

The likely source of these errors is the rate coefficients. The all-*trans* retinoic acid and the 13-*cis* isomer degrade to side-products faster than they interconvert. 13-*cis* also degrades much faster than all-*trans*. Under these circumstances, the mass-action ratio of

any two isomers never converges to the equilibrium constant, even when extrapolated to infinite time. In fact, the ratio by Urbach and Rando can be reproduced by using our rate constants when ignoring polymerization. The equilibrium percentage of 6.7 ± 1.1 , 61.4 ± 0.5 , 14.8 ± 0.6 and 17.6 ± 1.3 for 9, 13-*dicis*, *all-trans*, 9-*cis* and 13-*cis* consistent with our calculated values: 6.1, 61.4, 14.8 and 17.6 when we calculate at 300 minutes as they reported. The ratio will always represent a lower ratio and therefore a higher standard free energy for the faster-degrading isomer. This probably explains why both Rando's papers give 13-*cis* values that are too high, in one case much too high. Similarly, because 9-*cis* appears to be less susceptible to degradation than *all-trans*, earlier work underestimates its standard free energy.

Computations indicate the thermodynamic effect of the two isomerizations is largely additive, and our experimental data are not sufficiently precise to rule this out.

5.4 Conclusions

The standard free energies of the most stable isomers of retinoic acid will be of value to models examining the biosynthesis and catabolism of these important biological agents. Furthermore, the equilibrium constants and standard free energies are useful for studies of metabolism involving retinoic acid isomers. Most significantly, our results indicate 13-*cis* retinoic acid is markedly more stable than previously reported, probably due to an incorrect estimate of the degradation rate of the 13-*cis* isomer. The results also show that both single and double isomerization reactions follow first-order kinetics, suggesting a common intermediate for the acid-catalyzed isomerization which allows rotation about several formerly double bonds at once.

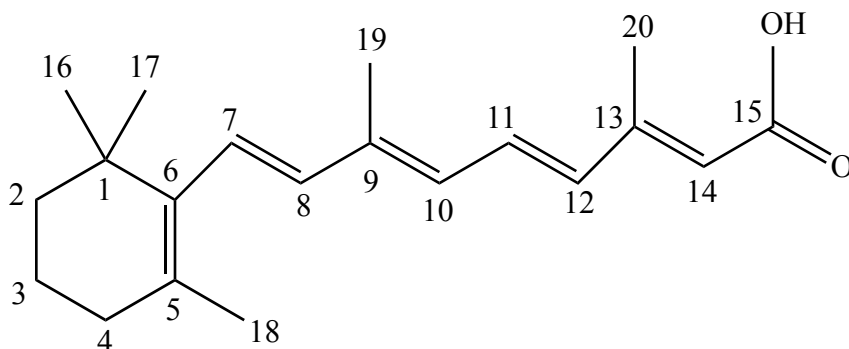
Computational results, using a moderately high level of theory combined with full thermochemistry and solvation corrections, reproduce experimental ΔG° value within 1 kJ/mol. This is a significant outcome for estimating the thermochemistry of isomers, such as 11-*cis* and 7-*cis*, which are less stable and only present at low concentrations.

References

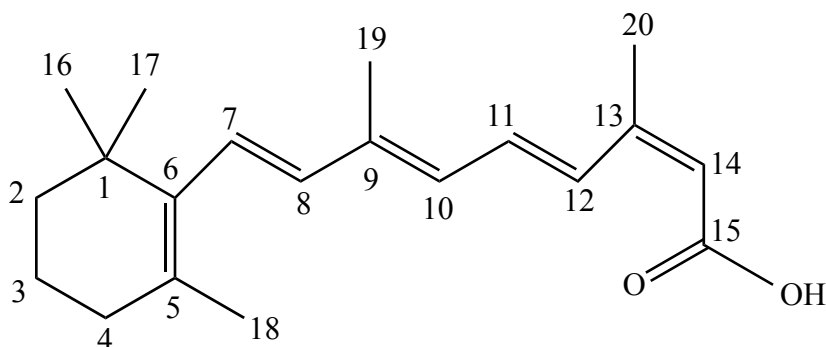
- [1] G. Wald, P. K. Brown, *Science* 1958, *127*, 222-226.
- [2] G. Wald, P. K. Brown, P. H. Smith, *Journal of General Physiology* 1955, *38*, 623-681.
- [3] aB. S. Chertow, W. S. Blaner, N. G. Baranetsky, W. I. Sivitz, M. B. Cordle, D. Thompson, P. Meda, *Journal of Clinical Investigation* 1987, *79*, 163-169; bG. Cabrera-Valladares, M. S. German, F. M. Matschinsky, J. H. Wang, C. Fernandez-Mejia, *Endocrinology* 1999, *140*, 3091-3096.
- [4] aK. Robertson, I. Mason, *Mechanisms of Development* 1995, *53*, 329-344; bJ. Briscoe, B. G. Novitch, *Philosophical Transactions of the Royal Society B-Biological Sciences* 2008, *363*, 57-70.
- [5] R. Hubbard, *Journal of Biological Chemistry* 1966, *241*, 1814-&.
- [6] R. R. Rando, A. Chang, *Journal of the American Chemical Society* 1983, *105*, 2879-2882.
- [7] J. Urbach, R. R. Rando, *Biochemical Journal* 1994, *299*, 459-465.
- [8] C. Lanvers, G. Hempel, G. Blaschke, J. Boos, *The FASEB Journal* 1998, *12*, 1627-1633.
- [9] H. Chen, M. R. Juchau, *Biochemical Journal* 1998, *336*, 223-226.
- [10] M.-N. Marchetti, E. Sampol, H. Bun, H. Scoma, B. Lacarelle, A. Durand, *Drug Metabolism and Disposition* 1997, *25*, 637-646.
- [11] J. O. Sass, G. Tzimas, H. Nau, *Drug Metabolism and Disposition* 1997, *25*, 1447.
- [12] J. Urbach, R. R. Rando, *Febs Letters* 1994, *351*, 429-432.
- [13] S. N. Hawk, M. A. Satre, *Faseb Journal* 2001, *15*, A960-A960.
- [14] M. Tsukada, M. Schroder, T. C. Roos, R. A. S. Chandraratna, U. Reichert, H. F. Merk, C. E. Orfanos, C. C. Zouboulis, *Journal of Investigative Dermatology* 2000, *115*, 321-327.
- [15] A. Murayama, T. Suzuki, M. Matsui, *Journal of Nutritional Science and Vitaminology* 1997, *43*, 167-176.
- [16] R. W. Curley, J. W. Fowble, *Photochemistry and Photobiology* 1988, *47*, 831-835.
- [17] S. R. Kunchala, T. Suzuki, A. Murayama, *Indian Journal of Biochemistry & Biophysics* 2000, *37*, 71-76.
- [18] R. M. McKenzie, D. M. Hellwege, M. L. McGregor, E. C. Nelson, *Lipids* 1979, *14*, 714-717.
- [19] A. K. Sakhi, T. E. Gundersen, S. M. Ulven, R. Blomhoff, E. Lundanes, *Journal of Chromatography A* 1998, *828*, 451-460.
- [20] R. S. H. Liu, A. E. Asato, *Tetrahedron* 1984, *40*, 1931-1969.
- [21] A. E. Derome, M. P. Williamson, *Journal of Magnetic Resonance (1969)* 1990, *88*, 177-185.
- [22] B. Perly, G. C. Pappalardo, M. Klaus, E. Montoneri, *Zeitschrift Fur Naturforschung Section B-a Journal of Chemical Sciences* 1988, *43*, 1072-1074.

Figure 5.1 Chemical structure of all-*trans*, 13-*cis*, 9-*cis*, 11-*cis*, 9,13-*dicis* retinoic acid

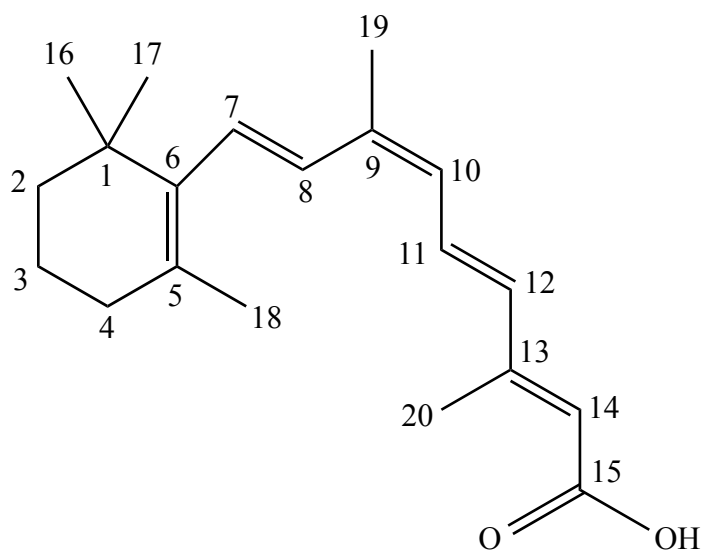
All-*trans* retinoic acid



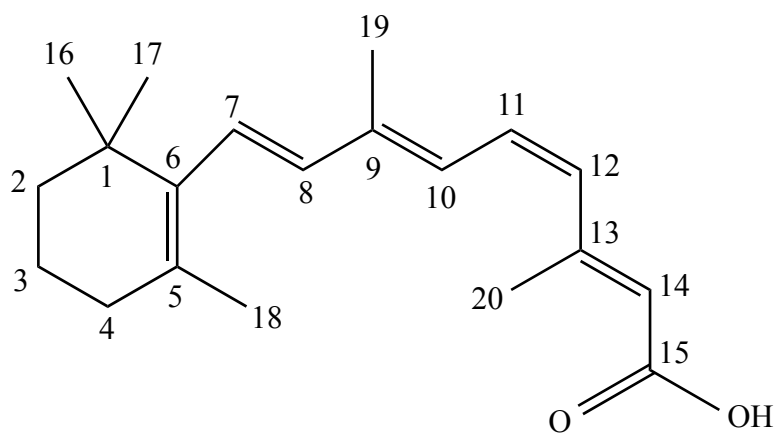
13-*cis* retinoic acid



9-*cis* retinoic acid



11-*cis* retinoic acid



9,13-dicis retinoic acid

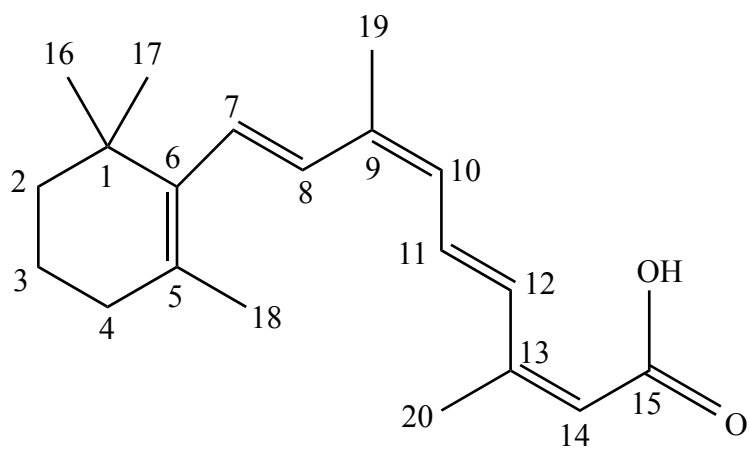


Figure 5.2 Kinetic scheme of trifluoroacetic acid catalyzed isomerization of retinoic acid in CD_2CL_2 . a being All-*trans*, b being 13-*cis*, c being 9-*cis* and d being 9,13-*dicis* retinoic acid. k_{aa} , k_{bb} , k_{cc} and k_{dd} are respectively the decay rate constants for all-*trans*, 13-*cis*, 9-*cis* and 9,13-*dicis* retinoic acid. k_{ab} , k_{ba} , k_{ac} , k_{ca} , k_{ad} , k_{da} , k_{bc} , k_{cb} , k_{bd} , k_{db} , k_{cd} , and k_{dc} represent isomerization rate constants.

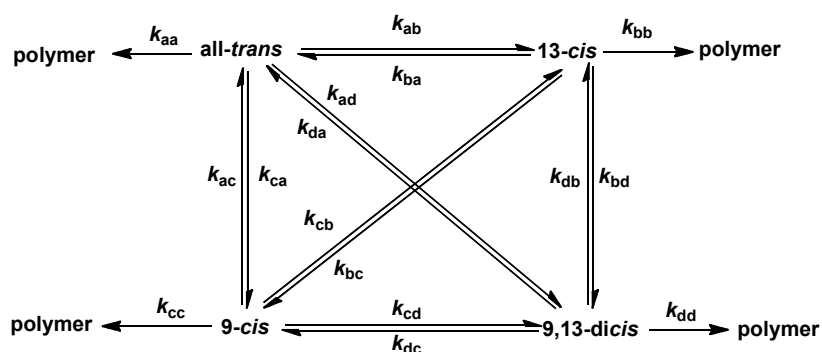


Figure 5.3 ^1H NMR spectrum of vinyl protons of all-*trans* retinoic acid

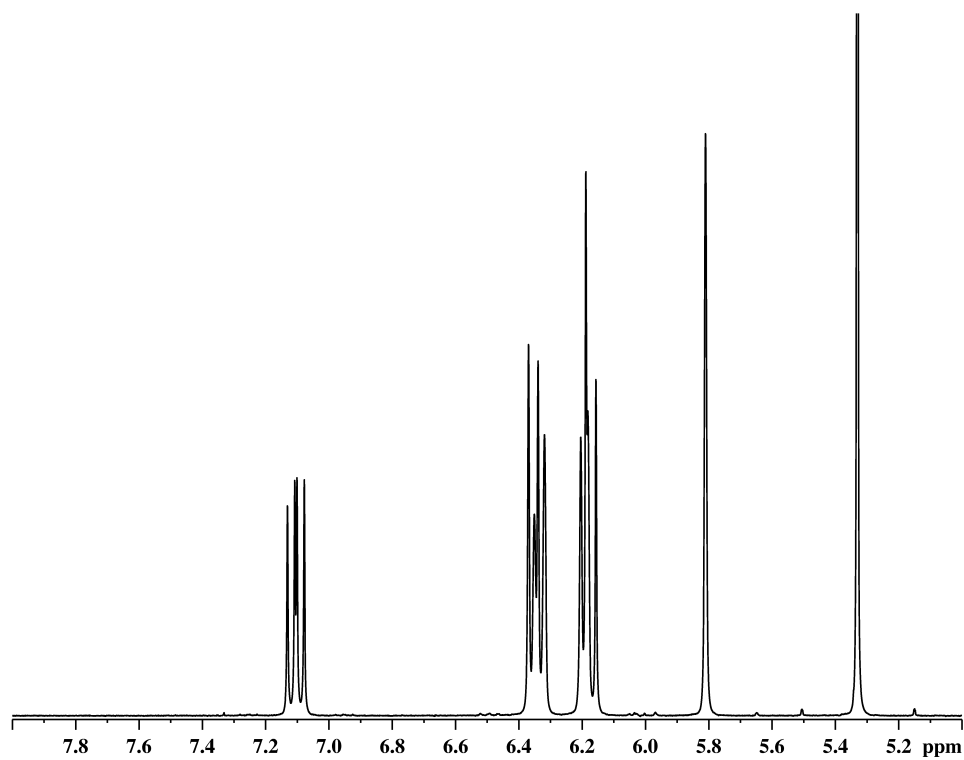


Figure 5.4 ^1H COSY NMR spectrum of vinyl protons of all-*trans* retinoic acid

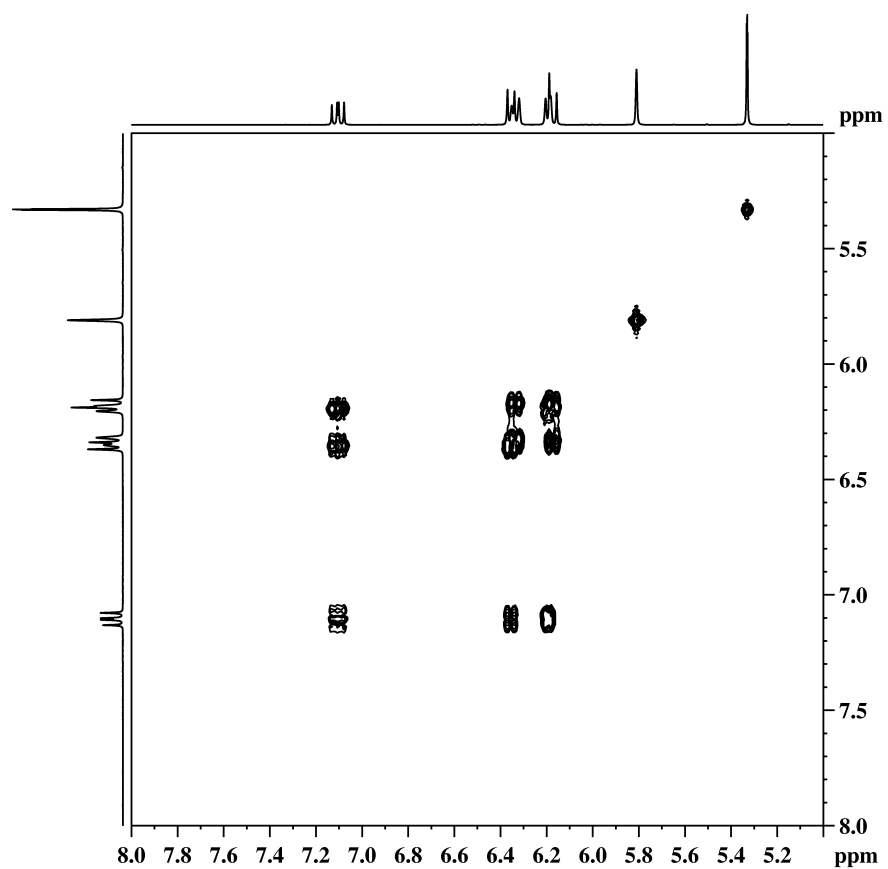


Figure 5.5 ^1H NMR spectrum of vinyl protons of 13-*cis* retinoic acid

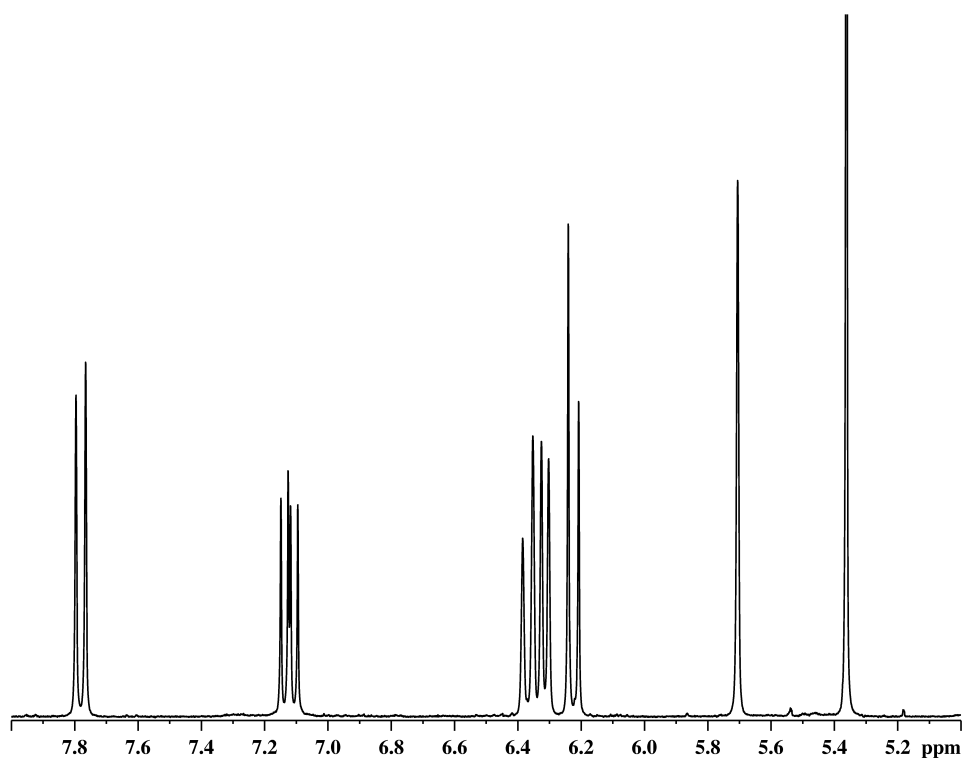


Figure 5.6 ^1H COSY NMR spectrum of vinyl protons of 13-*cis* retinoic acid

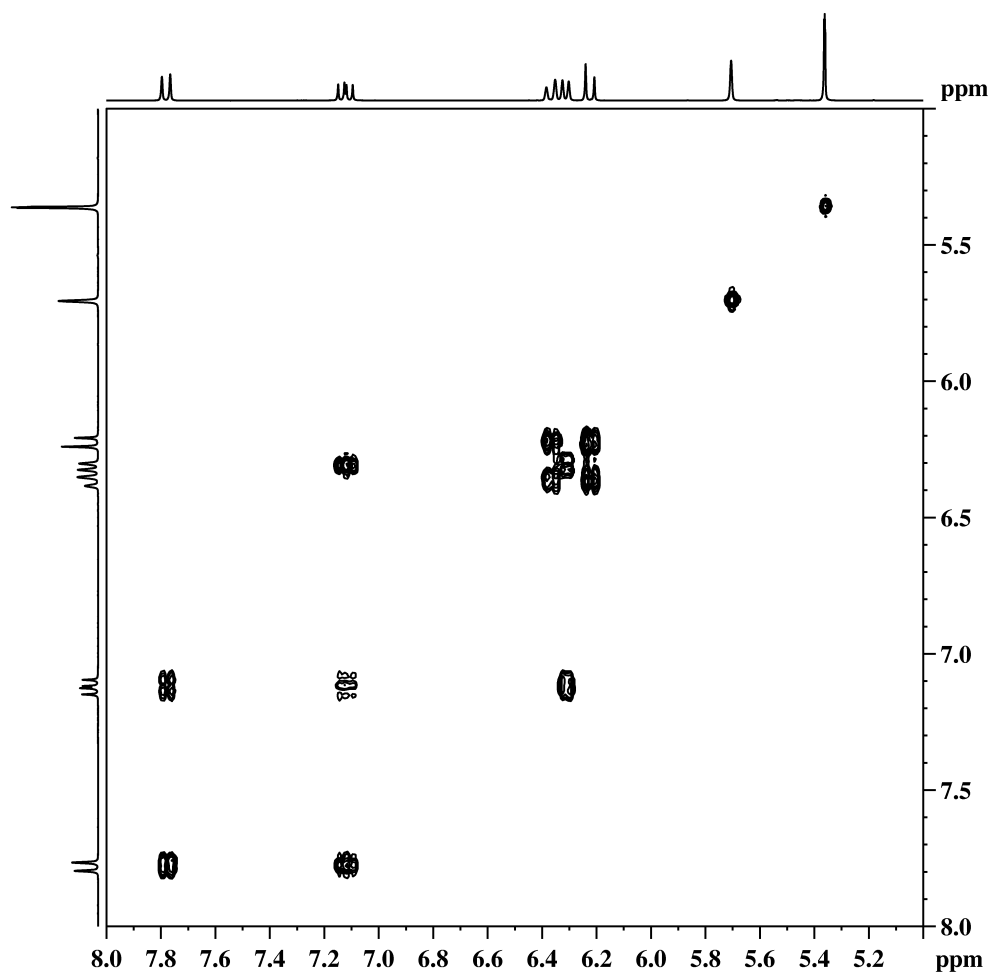


Figure 5.7 ^1H NMR spectrum of vinyl protons of 9-*cis* retinoic acid

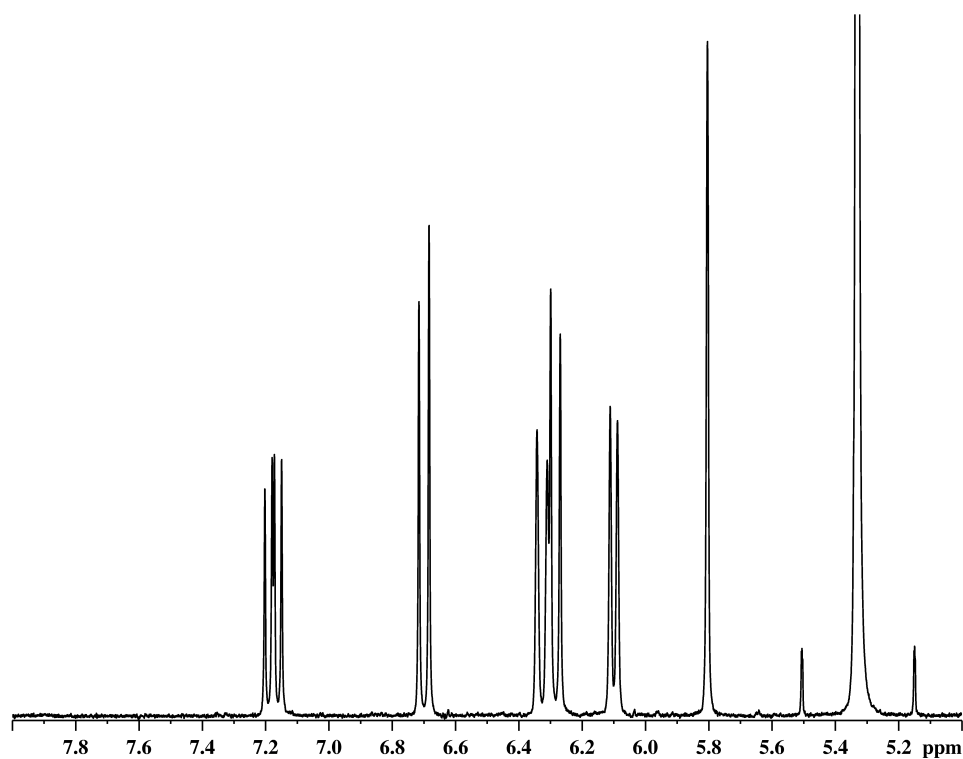


Figure 5.8 ^1H COSY NMR spectrum of vinyl protons of 9-*cis* retinoic acid

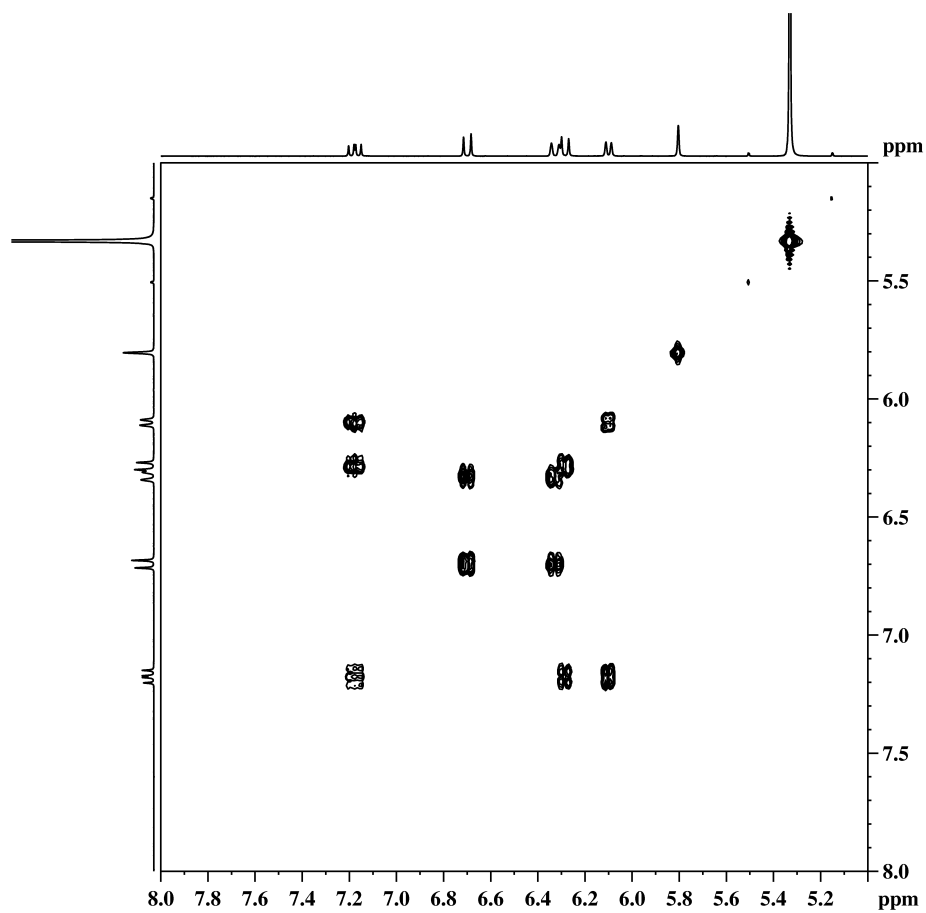


Figure 5.9 ^1H DQF-COSY NMR spectrum of vinyl protons of I_2 catalyzed 13-*cis* retinoic acid

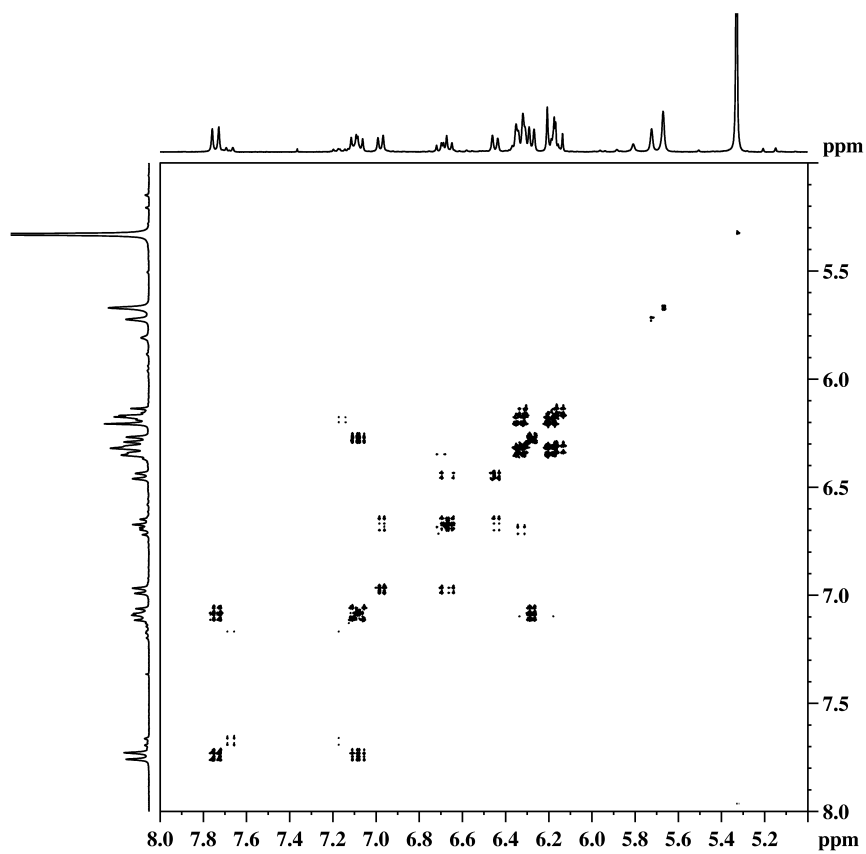
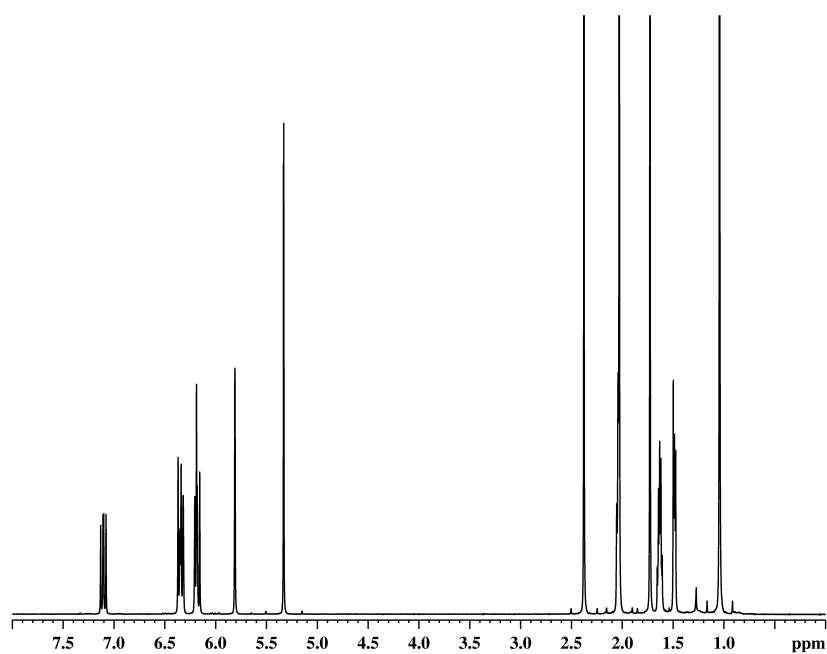
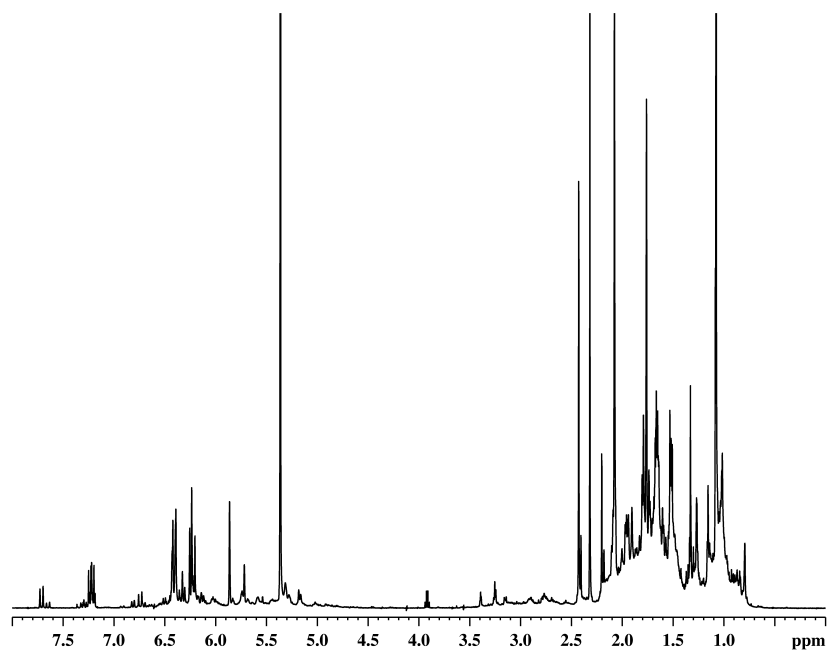


Figure 5.10 The first graph is all-*trans* retinoic acid ^1H NMR full spectrum; the second graph is trifluoroacetic acid catalyzed all-*trans* retinoic acid ^1H NMR spectrum after reacting 150 minutes; the third graph is trifluoroacetic acid catalyzed all-*trans* retinoic acid ^1H NMR spectrum after reacting 735 minute





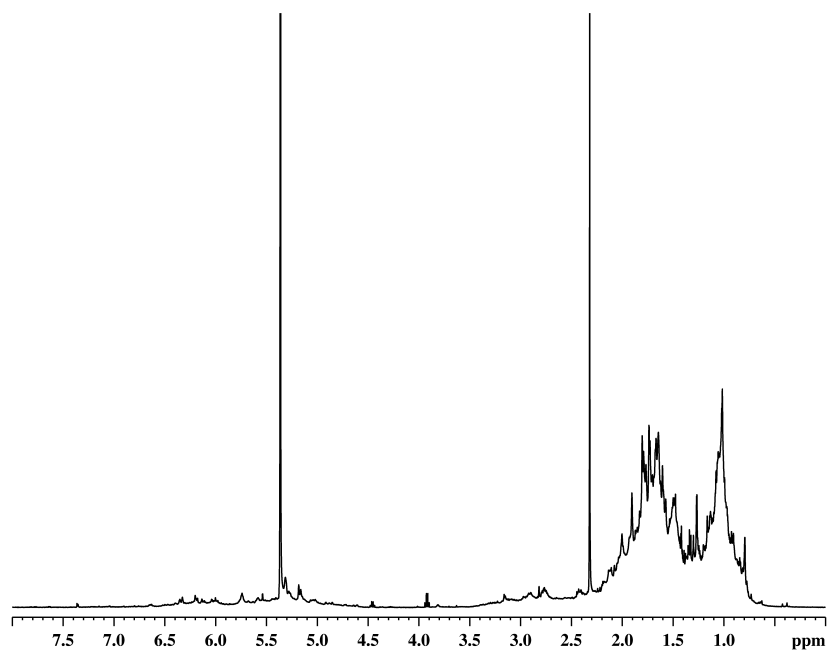


Figure 5.11 Mole fractions of retinoic acid isomers with respect to time after adding trifluoroacetic acid starting from all-*trans* retinoic acid. The dot is experimental data and the line is fitted graph to figure out rate constants.

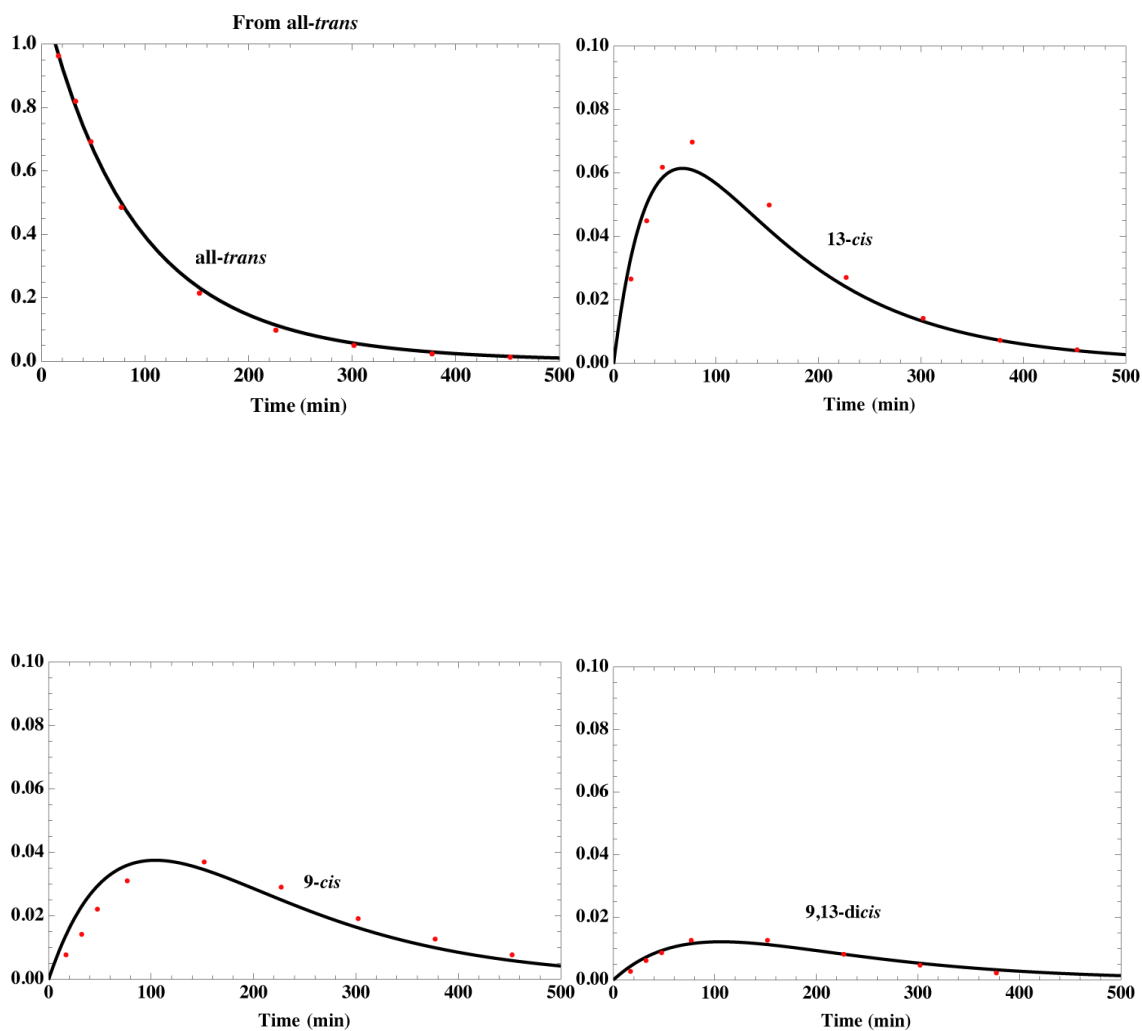


Figure 5.12 Mole fractions of retinoic acid isomers with respect to time after adding trifluoroacetic acid starting from 13-*cis* retinoic acid. The dot is experimental data and the line is fitted graph to figure out rate constants.

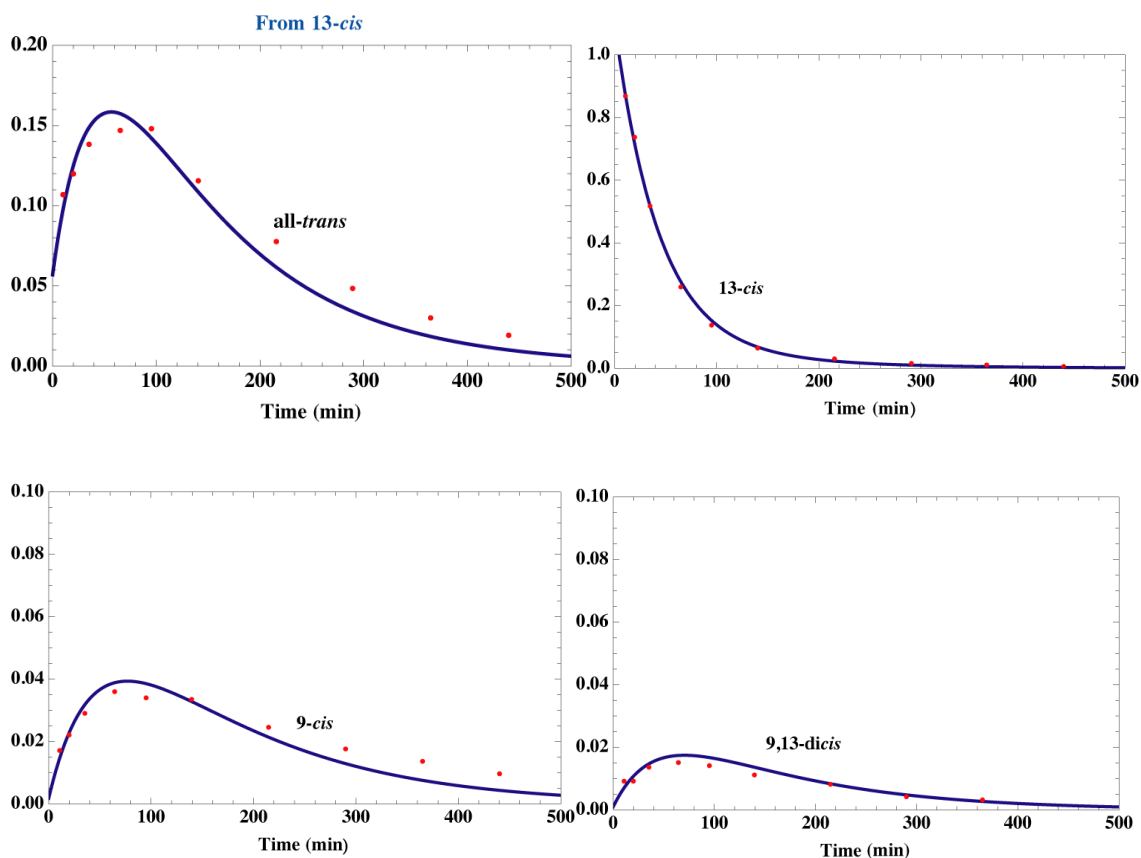


Figure 5.13 Mole fractions of retinoic acid isomers with respect to time after adding trifluoacetic acid starting from 9-*cis* retinoic acid. The dot is experimental data and the line is fitted graph to figure out rate constants.

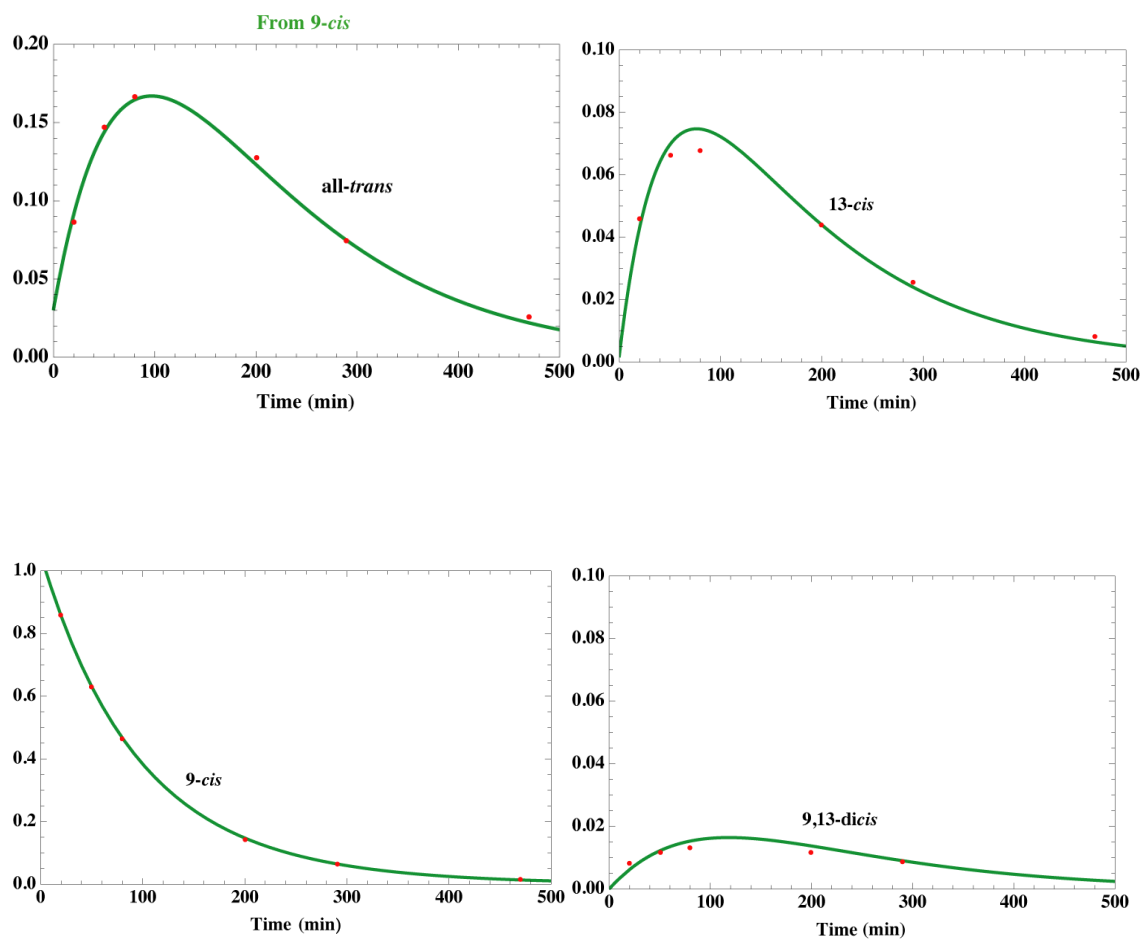


Table 5.1 Chemical shifts and coupling constants of vinyl protons of retinoic acid isomers.

isomer	Vinyl protons (ppm)						Coupling constant (Hz)		
	7	8	10	11	12	14	7,8	10,11	11,12
<i>All-trans</i>	6.172	6.35	6.193	7.10	6.354	5.810	16.08	11.45	14.96
				(7.192)					
<i>13-cis</i>	6.336	6.192	6.282	7.09	7.749	5.674	16.11	11.50	15.27
				(7.181)	(7.6803)				
<i>9-cis</i>	6.326	6.699	6.099	7.176	6.283	5.804	15.97	11.53	14.94
		(6.7088)		(7.2654)					
<i>9,13-dicis</i>				7.257	7.6162			11.01	15.18

Table 5.2 Rate constants of trifluoroacetic acid catalyzed interconversion of retinoic acid in CD₂Cl₂ detected by NMR spectra. The experimental rate constants with column molecules being reactants and row molecules being products.

	<i>all-trans</i> (d⁻¹)	<i>13-cis</i> (d⁻¹)	<i>9-cis</i> (d⁻¹)	<i>9,13-dicis</i> (d⁻¹)
<i>all-trans</i>	12.5	7.49	5.96	8.29
<i>13-cis</i>	3.22	22.2	3.54	7.27
<i>9-cis</i>	1.18	1.63	5.25	1.82
<i>9,13-dicis</i>	0.434	0.884	0.480	0

Table 5.3 Equilibrium constants of trifluoroacetic acid catalyzed retinoic acid isomerization in CD₂Cl₂

	13-<i>cis</i> (kJ/mol)	9-<i>cis</i> (kJ/mol)	9,13-<i>dicis</i> (kJ/mol)
Computed (this work)	0.81	3.84	5.32
Experimental (this work)	2.11 ± 0.10	4.01 ± 0.13	7.31 ± 0.7
Retinoic acid ^[7]	3.3	3.3	5.0
Retinal ^[6]	2.46	3.57	5.95

Chapter 6

Solid-State NMR of Amino Acids, and the Origin of Life

Abstract

This chapter reports the use of high-resolution solid state ^{13}C NMR, using cross polarization (CP) and magic angle spinning (MAS), to study chirality in the crystallization of aspartic acid. We show that, contrary to recent reports, DL-aspartic acid crystallizes over most of its temperature ranges as racemic crystals rather than a conglomerate of enantiomeric crystals, regardless of whether the solution was prepared by solution of the racemate or by mixing solutions of the pure enantiomers. Over virtually all the solution temperature range at 1 bar pressure, the racemic crystal of aspartic acid is thermodynamic stable, while its conglomerate is metastable. In contrast and in agreement with literature, glutamic acid crystallizes as a conglomerate under both kinetic and thermodynamic conditions. Solid-state NMR results are confirmed by powder X-ray diffraction (PXRD).

6.1 Introduction

Chirality, the category of optical isomerism that makes a molecule not superimposable on its mirror image, is a fundamental characteristic of life. Isomers that rotate light clockwise and anticlockwise are termed ‘dextrorotary’ (D) and ‘levorotary’ (L) respectively. Pairs of chiral optical isomers, or enantiomers, share physical and chemical properties, but differ in the direction of their rotation of plane-polarized light. In biological systems, proteins are formed from L-amino acid enantiomers while carbohydrates are composed of D-isomers. The puzzle of how life adopted a unique chirality has attracted scientists’ attention for over a 100 years, since Pasteur discovered the phenomenon of optical isomerism¹. Chirality may have evolved after the evolution of life, or before it. Theories that propose the latter are called abiotic and fall generally into two classes: (i) an interaction that discriminates between enantiomers led to the preferential production or catabolism of one isomer; (ii) a serendipitous separation of isomers, perhaps combined with a mechanisms for amplification of the enantiomeric excess of a mixture, led to a system in which a single enantiomer predominated. Such mechanism have been comprehensively reviewed by Bonner², and will not be further examined here.

Crystallization is one possible process that might lead to an enantiomeric excess. Chiral molecules in a racemic solution can generally crystallize in one of two ways. Crystals of distinct enantiomers can be formed, excluding the other enantiomer from the growing crystal. The result of such crystallization from a racemic solution is what is called a conglomerate, which contains equal proportions by mass of individual crystallites, each containing a single enantiomer, either D or L. Alternatively, D and L isomers can

crystallize together, in equal proportions, in a racemic crystal, which will generally possess a mirror plane or center of inversion. Early experimental studies of such processes led to the formulation of Wallach's rule, which holds that racemic crystals, with both right-handed and left-handed molecules present, tend to find denser packing motifs, and therefore more favorable interactions. All 230 point groups are available to enantiomers, whereas only 65 have the mirror plane or inversion center that would permit a point group to accommodate a racemate without disorder³ Brock *et al.*⁴, examined 129 pairs of crystal structures where both racemate and enantiomeric crystal were known and concluded that racemic crystals in general indeed tended to be more stable, Although it has been pointed out that amino acid racemates do not typically have higher densities than the corresponding conglomerates, amino acids nonetheless tend to follow this rule³. While it is a possibility that a racemate could crystallize in a mixture of right-handed and left-handed crystals in a chiral group if D and L molecules in the crystal are not related by a symmetry element, this contingency has not to our knowledge ever been observed.

Serendipitous nucleation of a single enantiomeric crystal from a racemic solution, combined with a racemization reaction that replaces the depleted enantiomer, could clearly lead to spontaneous resolution of a solution (driven, thermodynamically, by whatever heat source drives evaporation). Such spontaneous resolution was first reported in 1952 for tri-o-thymotide⁵, but has not to our knowledge yet been observed for natural amino acids.

Our interest in this problem was driven by a remarkable paper from 2010, titled "The Origin of Life and the Crystallization of Aspartic Acid in Water" ⁶, which received considerable play in the popular science literature. The authors claimed at a solution of

DL aspartic acid produced by dissolving the racemic acid was different from a solution obtained by mixing equal proportions of D aspartic acid and L aspartic acid. The latter solution, the authors claimed, formed a 'conglomerate solution' that retained for some hours a memory of its formation. The authors hypothesized that the enantiomer molecules form long-lived intermolecular complexes, even in solution, that last "36 h at 25 C and 5 h at 45 C". While the 'conglomerate solution' persisted, a conglomerate solid, instead of the normal racemic solid, was obtained by crystallization of the solution. The authors also claimed even this slow rate of transformation could be further inhibited by addition of succinic acid (which is achiral). The authors suggested that these long-lived solutions might have led to preferential crystallization of enantiomers in the early earth, possibly explaining the chirality of life.

Our immediate reaction to this paper was skepticism bordering on outright disbelief. The idea of long-lived mono-enantiomeric regions in amino acid solutions in water goes against what we know of the statistical mechanics of aqueous solutions, in which complexes last on the order of nanoseconds, and solvation by water is competitive with solvation by other amino acid molecules. It also runs directly contrary to 298 K diffusion constant measurements⁷ for aspartic acid, which lie in the range of $0.7 - 0.8 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ and are only weakly dependent on concentration, which are typical for a small molecule in aqueous solution. The consensus of the field is that activity coefficients of amino acids in dilute aqueous solution are approximately unity⁸. Nonetheless, because the results of Lee and Lin were quite compelling, we resolved to try to repeat the work, using solid-state NMR as a probe of the crystalline phase formed by crystallization of aspartic acid solutions.

A saturated solution is one where the free energy of the dissolved molecule equals that of the crystal. In a racemic solution, this becomes slightly more complicated, because there are two solids to consider: the racemic, and the enantiomeric crystal. Thus, in principal a racemic solution has solubilities both with respect to the conglomerate and racemate solid; one of these will be metastable rather than a true equilibrium solubility. But, the solubility measurements may be complicated by not knowing which solid is present.

Early room temperature solubility data was collected by Dunn *et al*⁹. for both L- and DL-aspartic acid. Their van't Hoff plots for the racemic amino acid were quite curved, indicating a large heat capacity change. More recent measurements¹⁰ conform their observations, and were used in this work.

Crystal structures of both the enantiomeric and racemic crystals have been well-characterized. An early study of twinned L-aspartic acid revealed it to crystallize in non-centrosymmetric (chiral) group $P2_1$, with a single zwitterionic molecule in the unit cell, and a packing which contains only intermolecular hydrogen bonds¹¹. A more recent study¹² confirmed the essentials of the structure, with improved accuracy.

The structure of DL-aspartic acid was determined at about the same time¹³. The molecule, in the centrosymmetric group $C2/c$, has the expected mirror pair of enantiomers, which should have identical NMR chemical shifts. A 20 K structure of the DL-aspartic acid crystal¹⁴ explored the deformation electron densities in more detail and better defined the hydrogen bonds. A neutron diffraction study is also available¹⁵. A recent claim of another structure¹⁶ with a slightly different a axis length but essentially the same atomic positions and space group is likely just the same structure.

Two crystal structures of L-glutamic acid are known, both orthorhombic in the space group $P2_12_12_1$, a metastable¹⁷ ‘alpha’ form^{18,19} and a stable beta form²⁰. DL-glutamate crystallizes anhydrously from ethanol as a racemate²¹. In contrast to most amino-acids, glutamic acid tends to crystallize from water under normal conditions as a conglomerate²¹; its enantiomeric forms appear to be much less soluble than its racemate⁸.

It is almost a truism that NMR, like most other physical measurements, cannot determine the chirality of a molecule. In solution, D and L isomers can only be distinguished by NMR in a solution that is itself chiral — for example, one which contains a chirally coordinated paramagnetic shift reagent²². However, in the solid-state, the spectra of L and DL crystals are generally quite different. This is not a direct result of the chirality, but rather from a difference of packing and therefore intermolecular interactions between enantiomeric and racemic crystals. Thus, Diaz *et al.*²³ were able to show that the spectra of crystalline L-methionine and DL-methionine were different. While both crystal forms have only one chemically-distinct molecule in the unit cell, differences in intermolecular interactions lead to distinct cross-polarization/magic-angle spinning (CP-MAS) spectra. We had every expectation that the same method could be applied to aspartic acid.

6.2 Theory

Thermodynamics of racemic and conglomerate solids in equilibrium with solution

The solubility of aspartic acid is fairly low, and because it is a weak acid, it is largely zwitterionic in solution. We therefore set activity coefficients to 1 and assume molalities equal activities. The results of Chen *et al.*²⁴ bear out this assumption; at mole fractions

below 0.005 (correspond to molalities under 0.25 m) the activity coefficients of several common amino acids are effectively one.

We start with two solubility equilibria. K_c is the solubility product for the enantiomer or conglomerate and is simply the molality of a saturated solution of the pure enantiomer. K_r is the solubility product of the racemate, and is the product of molalities of the two enantiomers in a solution of the racemate, saturated with respect to the racemic solid (which may or may not be supersaturated with respect to the enantiomer). m_D and m_L are the molalities of the two respective enantiomers. We can now write m , the total molality, as $m_D + m_L$, and f , the fraction of L enantiomer, as $m_L/(m_D + m_L)$. Therefore, $m_L = fm$ and $m_D = m - m_L = m(1-f)$. The molality of the conglomerate saturated solution is now $m = K_c/f$ for $0.5 < f < 1.0$ and $m = K_c/(1-f)$ for $0 < f < 0.5$.

For the racemate, we have for the saturated solution

$$K_r = m_D m_L = m^2 f(1-f).$$

The solubility of the conglomerate is clearly highest at $f = 0.5$, where $m = 2K_c$. At this value of f , the solubility of the racemate is $m = 2\sqrt{K_r}$. Obviously, if $\sqrt{K_r}$ is greater than K_c at this molality, the racemate will never precipitate (at least, under thermodynamic conditions). If $\sqrt{K_r} < K_c$, there will be a region of f where the racemate precipitates, although as $f \rightarrow 0$ or $f \rightarrow 1$, pure enantiomeric crystals will appear. This is shown in Figure. 1 for $K_c = 1$ and $\sqrt{K_r} = 0.5, 0.75$, and 1.0 ; the figure was computed from the equations above. As can be seen, for the lower values of K_r , there exists a region at intermediate values of f where the solution is in equilibrium with the solid racemate. At higher values of K_r , the racemic solid is never thermodynamically stable. Obviously, by

the Gibbs phase rule, solution, racemic solid, and conglomerate can only coexist at equilibrium at a triple point of temperature and molality.

Temperature dependence of K_r and K_c

The extended van't Hoff equation gives the temperature dependence of the saturated solubility

$$\ln m_L = \frac{1}{R} \left(a_L + \frac{b_L}{T} + c_L \ln T \right) \quad (1)$$

with $c_L = \Delta_{\text{soln,L}} C_p^\circ$

$$b_L = -\Delta_{\text{soln,L}} H^\circ + \Delta_{\text{soln,L}} C_p^\circ T^\circ$$

$$a_L = \Delta_{\text{soln,L}} S^\circ - \Delta_{\text{soln,L}} C_p^\circ + \Delta_{\text{soln,L}} C_p^\circ \ln T^\circ$$

$$\ln m_{\text{DL}} = \frac{1}{R} \left(a_{\text{DL}} + \frac{b_{\text{DL}}}{T} + c_{\text{DL}} \ln T \right) \quad (2)$$

Note, however, that the molality of L aspartate in a solution of DL aspartate is $m_{\text{DL}}/2$.

Solubility data for both enantiomer and racemate were collected by Apelblat and Manzurola¹⁰. These were fit to the equations above, and the fits are compared with experimental data in Figure 2.

	a (kJ mol ⁻¹)	b (J mol ⁻¹ K ⁻¹)	c (J mol ⁻¹ K ⁻¹)
L-aspartate	-42.0	-1509	236
DL-aspartate	22.1	-43	16

Table 1: fits to the solubilities of L-aspartic and DL-aspartic acid, from the data given in 10, using eqns. 1 and 2.

Thermodynamic parameters are tabulated in Table 1, assuming a 1 mol kg⁻¹ standard state.

Although these parameters vary somewhat from the results of Chen *et al*²⁴, who used Fasman's earlier solubilities, they are in reasonably close agreement.

If we plot $m_{DL}/2$ and m_L versus temperature (Figure 3), the fits predict that the racemate and conglomerate solids will both be at equilibrium with a saturated solution of the racemate at 335.8 K. Above that temperature, the conglomerate solid will be more stable; below it, the racemate. We can also plot phase diagrams as a function of temperature. These are shown in Figure 4. Again, at low temperature, we predict crystallization of the racemate; at higher temperature, the conglomerate.

6.3 Materials and Methods

Samples of L-, D-, DL-aspartic acid and D- and L-glutamic acid were purchased commercially from Sigma chemical company. DL-glutamic acid was made from equimolar L- and D-glutamic acid.

Sample preparation

(1) To check on the 'solution memory' finding, stock solutions of L- and D-aspartic acid were made up by dissolving 0.159 g of pure material in 20 mL of water to give a concentration of 0.0597 *m*, the reported saturated concentration of the enantiomeric solutions at 35°C. The solutions were constantly stirred and heated at 60 °C during 2 hours to ensure complete dissolution of the solute. No crystallization was observed at room temperature. 10 mL D-aspartic acid and 10 mL L-aspartic acid were then mixed at room temperature, aliquots being removed at 1 m, 20 m, 40 m and 1 h, respectively. Aliquots were promptly placed in a vacuum dessicator and evaporated rapidly; freezing was observed within 30 seconds.

(2) To check if the solutions could be seeded with enantiomers, 0.3 g DL-aspartic acid was fully dissolved in 30 mL room temperature distilled water solution at 35°C, cooled to room temperature and then 2 mg solid L-aspartic acid was added. The sample was rapidly evaporated by the method described in (1).

(3) To check if enantiomeric crystals could be produced at elevated temperatures, crystalline samples were prepared from 0.3 g DL-aspartic acid fully dissolved in 30 mL distilled water at 35°C. Aspartic acid crystals were produced by evaporation at 60, 71, 81

and 91 °C, over time intervals from 12 h (60°C) to 2.5 h (91°C). Two samples were prepared; the first using evaporation in an open beaker, which took approximately 2.5 hours; the second using a beaker almost fully covered with a watch-glass, which took approximately 50 hours to evaporate.

For comparison with a material previously reported to crystallize as a conglomerate, a mixture of 0.3 g L-glutamic acid and 0.3 g D-glutamic acid were fully dissolved in 25 mL distilled water at 60 °C. Again, half of the solution was evaporated fast at 25°C (12 hours), in an open evaporation dish of 10.5 cm diameter, located inside a hood. The other half was evaporated more slowly, from a 50 mL beaker on the lab bench over a period of 2-3 days.

¹³C CP-MAS NMR Measurements

¹³C CP-MAS NMR spectra were obtained with a Bruker AVANCE 600 MHz spectrometer at field strength of 14.1 T (150.91 MHz ¹³C). Powder samples, crushed as finely as possible by using a mortar and pestle, were spun at 8 kHz in a 4 mm cylindrical rotor inside a Bruker ¹H/BB CP-MAS probe head. ¹³C chemical shifts were calibrated indirectly against the adamantane methylene peak of 28.8 ppm. The proton 90° high power pulse, contact time and recovery delay time were 4.5 μs, 1 ms and 240 s, respectively. Very long delay times were needed since the ¹H relaxation time in aspartic acid is quite long, presumably due to strong hydrogen bonding restricting the –NH₃⁺ group rotation²⁵. 256 scans were accumulated to obtain a satisfactory signal-to-noise ratio. All spectra of aspartic acid were analyzed using identical instrumental conditions and data processing procedures. Typical data collection time per sample was 17 hours.

Powder X-ray Diffraction (PXRD)

The PXRD spectra of powder samples were collected on a Rigaku D/Max-B X-ray diffractometer using Cu K α radiation (1.542 Å). The polychromatic beam with ~1 deg divergence and about 10 mm height was incident on the sample and the scattered beam was filtered through a graphite crystal monochromator. A 0.8 mm slit was inserted in front of the detector. A scintillation counter detector was used. The scanning rate was set at $2\theta = 0.05^\circ/\text{s}$ over a range from 10° to 35° .

6.4 Results and Discussion

The ^{13}C CP-MAS NMR spectra of commercial crystalline DL-Asp, L-Asp and an equimolar mixture of solid L- and DL-Asp are shown in Figure 5. The chemical shifts are shown in the figure. These spectra show that the CP-MAS NMR resonances of DL and L are easily distinguishable even in a mixture of both, with only one of the carbonyl peaks overlapped. In particular, the methylene (β -carbon) and methine (α -carbon) peaks of these two crystals are completely separated. We confirmed also that L-aspartic acid and D-aspartic acid give indistinguishable spectra.

The spectrum on the bottom right of Figure 5 was obtained by fast evaporation of the rapidly evaporated 1 minute mixture of L- and D- aspartic acid as described in section (1) of the methods. Contrary to the assertions of Lee and Lin⁶, the spectrum is essentially identical to that of DL-aspartic acid, and is not that of a conglomerate. Thus, the solid material, crystallized rapidly one minute after mixing solutions of L-aspartic acid and D

aspartic acid, is solid racemic DL-aspartate. The solutions retain no 'memory' for even this short length of time. Samples with longer mixing times gave similar spectra, with no evidence of conglomerate ever being formed.

We also obtained powder X-ray diffraction (PXRD) of rapidly crystallized aspartic acid mixtures (Figure 6). Again, the enantiomeric L-aspartic acid crystal gives a diffractogram quite different from racemic DL-aspartic acid; the diffractogram of a sample from rapid crystallization of the mixture of enantiomeric solutions showed only reflections from the racemate (although the broad baseline may indicate rapid crystallization also formed some amorphous material). Indeed crystallization under case (1) and (2) sample preparation conditions gave without exception purely racemic DL-aspartic acid. ^{13}C CP-MAS spectra indicate that even with seeding of racemic solutions at room temperature with enantiomeric crystals, crystallization of conglomerates did not occur. This result was also confirmed by PXRD.

Since theoretical analysis of the solubility data indicated that conglomerate formation is actually more likely at high temperature rather than low temperature, we turned to solutions of DL-aspartic acid crystallized at considerably higher temperatures. At 60° C, 70°C and 80°C, only racemate solid was obtained. However, fast crystallization at 91°C gave the first evidence of coexistence of a mixture of racemate and conglomerate (Figure 7). Only racemate was produced when repeated with slower crystallization. Thus, conglomerate seems to be produced from solutions of DL-aspartate only at high temperatures, and then only under kinetic rather than thermodynamic conditions.

Finally, for comparison, we examined the results of rapidly crystallizing solutions of DL-glutamic acid, where previous work suggests that the conglomerate is stable. Commercial

L-glutamic acid gives the spectrum shown in Figure 8 (top). Only a single crystalline form in vthe figure shows either the α or the β phase of L-glutamic acid existing. Rapid evaporation of a DL-glutamic acid solution gave the middle spectrum. The spectrum shows that a second phase is present in addition to the main conglomerate phase. The bottom spectrum obtained by slow evaporation indicate that different minor phase is observed. It is likely that one of the two minor phases is the β crystalline form of enantiomeric glutamic acid, while the other is the racemate; however, unambiguous identification awaits future research.

This work strongly suggests that the conclusions of Lee and Lin⁶ are erroneous. Regardless of the time solutions of L-aspartic acid and D-aspartic acid are allowed to mix, evaporation at room or lower temperature leads purely to racemic solid DL-aspartic acid. We would be tempted to suggest that the previous authors had failed to ensure the absence of enantiomeric seed crystals in their mixtures, which might lead to crystallization of a thermodynamically metastable conglomerate. However, we were ourselves unable to successfully seed the conglomerate at low temperature, and so this possibility is speculative at best. Otherwise, the previous authors' results remain inexplicable; they disagree with known diffusion constants of aspartic acid in aqueous solution, with the activity constants of amino-acids, which rule out large-scale, long-lived complex formation between amino acids, and with expectations from statistical thermodynamics. Our data also indicate the higher-temperature solubility data for L-aspartic acid and DL-aspartic acid may be somewhat in error; we find that racemate crystals are produced only above 90°C, and apparently under only kinetic conditions, whereas published data would indicate thermodynamically preferential crystallization of

the conglomerate at 60°C. It is perhaps notable that published experimental data above 60°C¹⁰ appear somewhat noisier, and their deviations from the fit curves are larger. It may be pertinent to note that under conditions where the free energy of racemate and enantiomer crystals are very close, metastable crystallization is more likely, and it may be unwise to assume a solution of any DL-aspartic acid is saturated with respect to the thermodynamically stable polymorph without examining the identity and composition of the solid present. The solubility of a DL-amino acid solution in this region where conglomerates and racemates are close to equilibrium requires great care to ensure the racemate and not the conglomerate is the actual solid phase present.

6.5 Conclusions

We find that the chiral D or L, and the racemic DL crystal forms of aspartic acid, have distinctly different ¹³C CP-MAS NMR spectra, and we can therefore use this technique to determine the proportions of chiral and racemic crystals in a polycrystalline solid. Phase diagrams indicate that aspartic acid conglomerate can only thermodynamically be formed in high temperature. Our experiment shows that fast evaporation of aspartic acid solution can give a mixture of conglomerate and racemate above 91 °C, while slow evaporation only gives racemate. Even extremely rapid crystallization does not give enantiomeric crystals at lower temperatures and, as expected, racemic solutions formed by mixing chiral solutions are indistinguishable from racemic solutions obtained by dissolving racemic crystals below 91 °C. Our results were confirmed by powder X-ray diffraction.

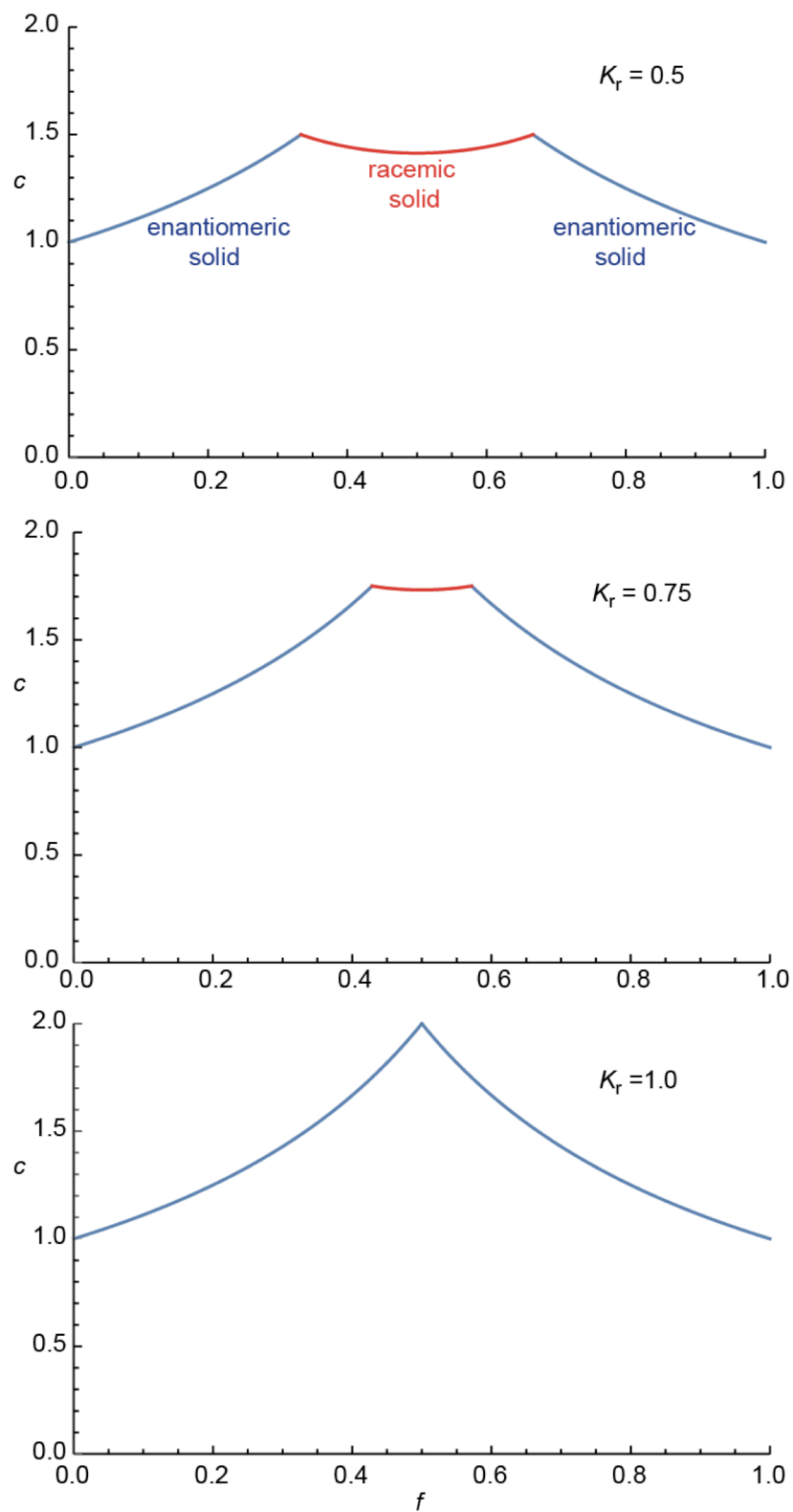


Figure 6.1 Theoretical phase diagrams for mixtures of L and D enantiomers, with a enantiomer solubility product of $K_c = 1$, at various values of the racemate solubility product K_r . c and f are defined in the text.

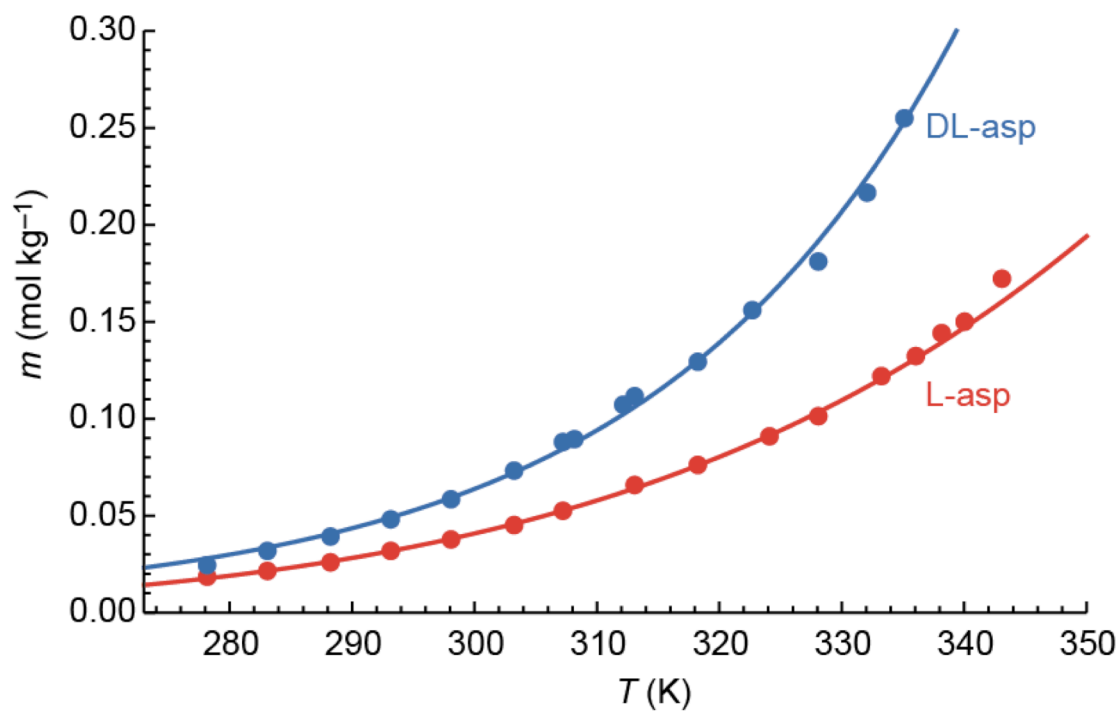


Figure 6.2 (Points) Solubilities of L-aspartic acid and DL-aspartic acid, from reference 10. (Lines) Fits to eqns. 1 and 2.

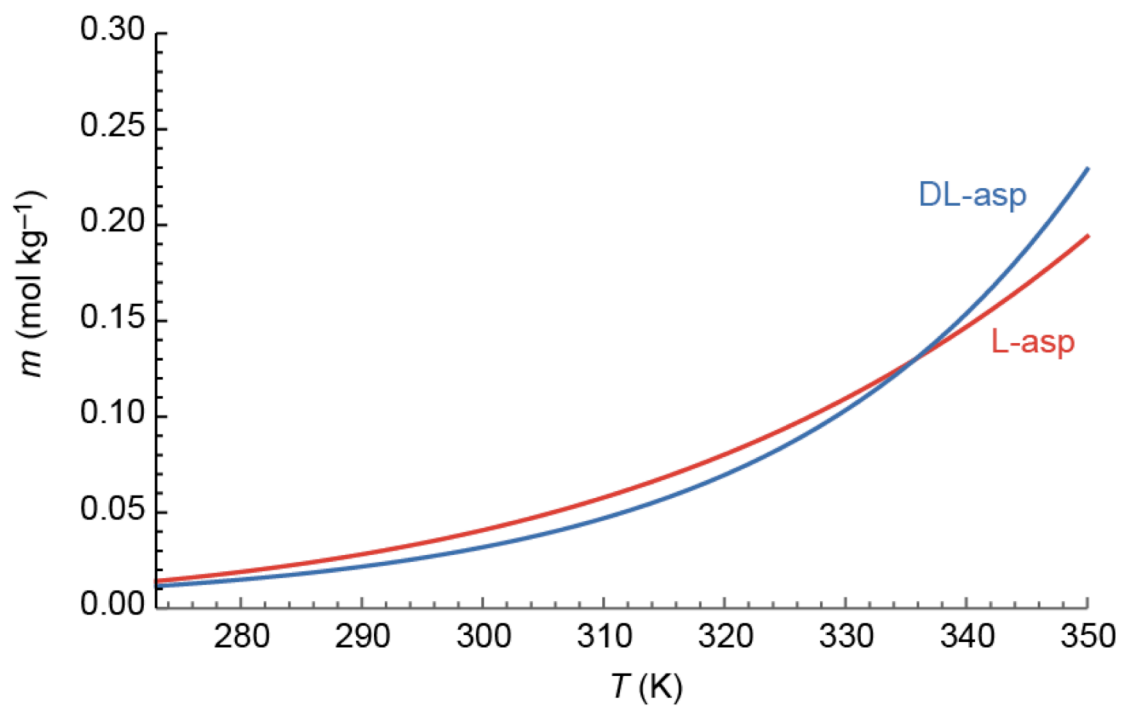


Figure 6.3 (Red) The fit solubility of L-aspartic acid in a DL aspartic acid solution, with respect to L-aspartic acid solid, compared with (blue) the solubility of L-aspartic acid in a DL aspartic acid solution, with respect to DL-aspartic acid solid.

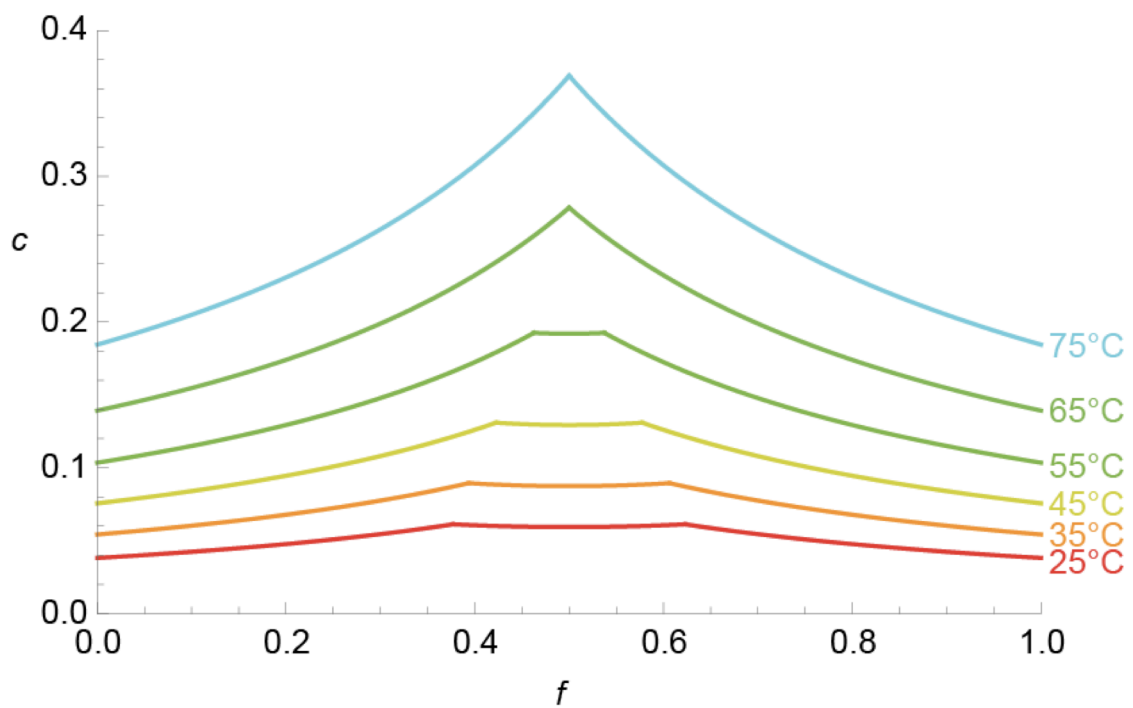


Figure 6.4 Phase diagram for crystallization of mixtures of L and D aspartic acid, using temperature dependent solubilities as a function of temperature depicted in Figure 6.2.

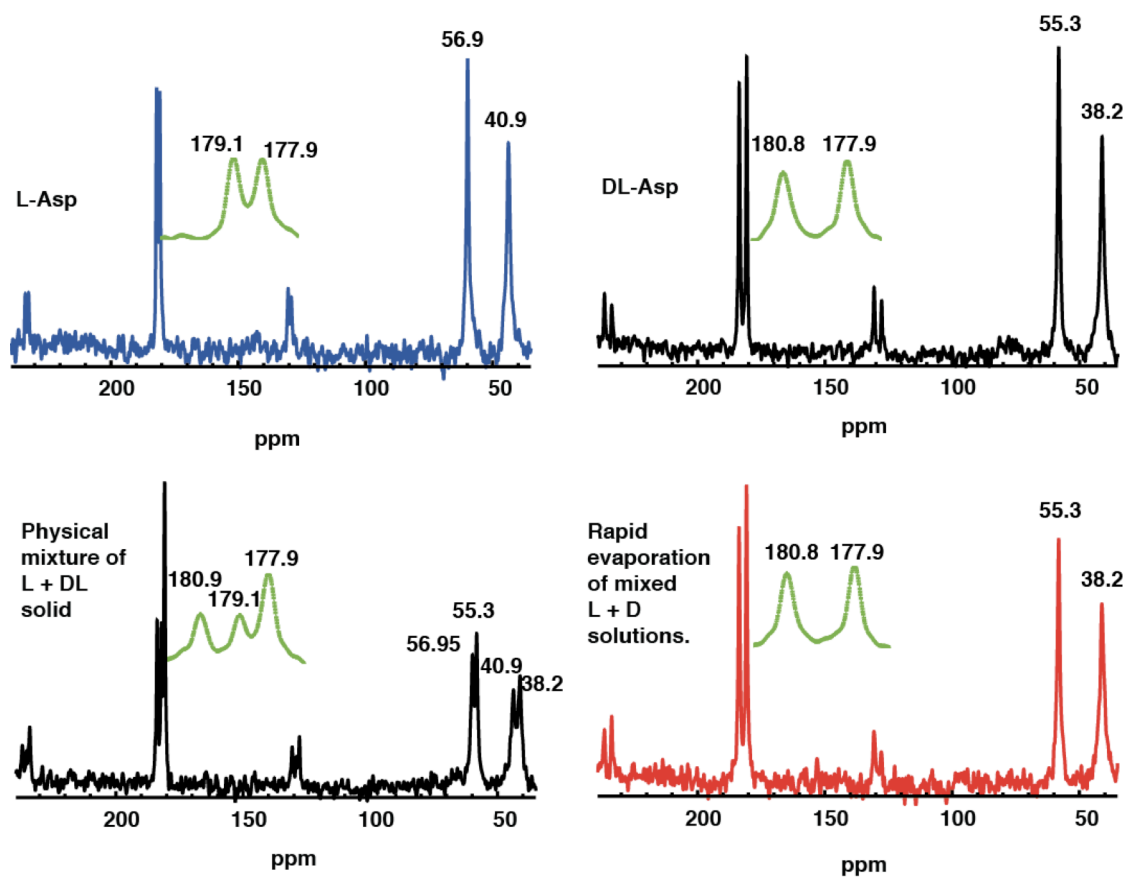


Figure 6.5 (Top) CP-MAS ^{13}C NMR spectra of commercial L-aspartic and DL-aspartic acid. (Bottom left) spectrum of an equimolar physical mixture of the L- and DL aspartic acid crystals (Bottom right) spectrum of crystals from a rapidly evaporated mixture of solutions of D- and L-aspartic acid.

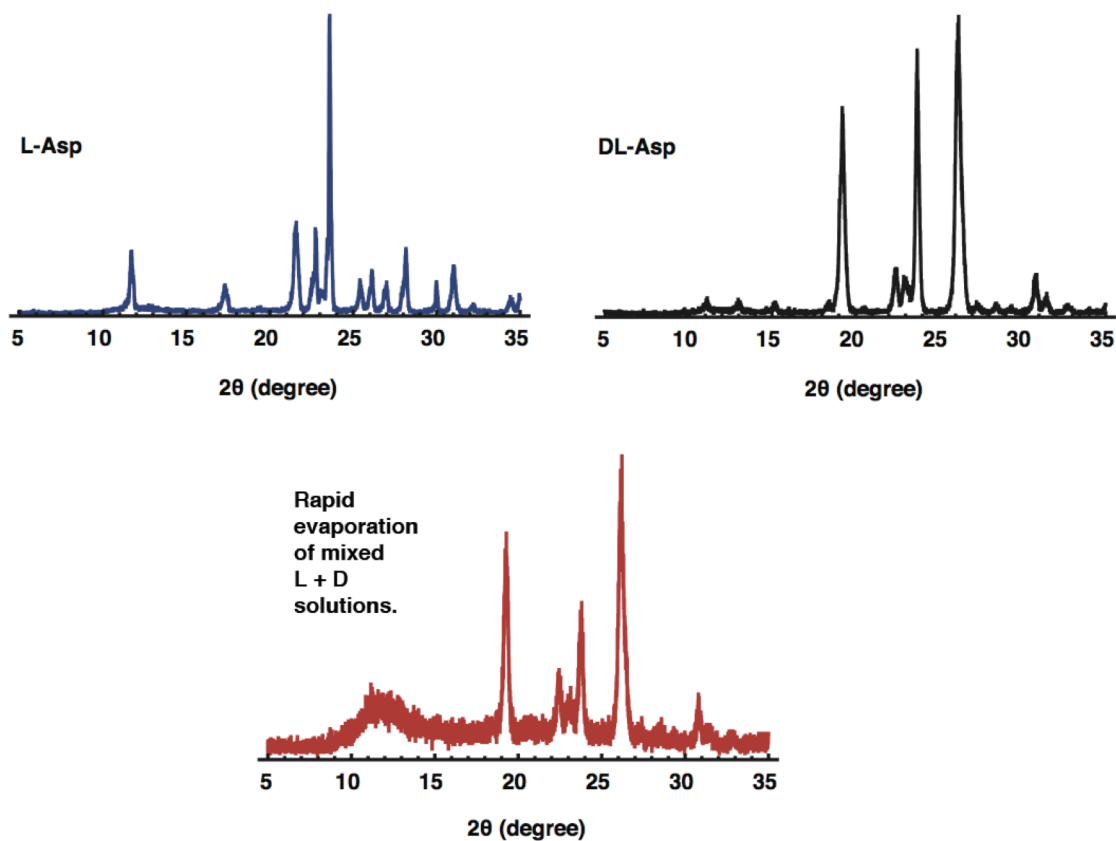


Figure 6.6 (Top) Powder X-ray diffraction of commercial L-aspartic and DL-aspartic acid. (Bottom) Powder X-ray diffraction of crystals from a rapidly evaporated mixture of solutions of D- and L-aspartic acid.

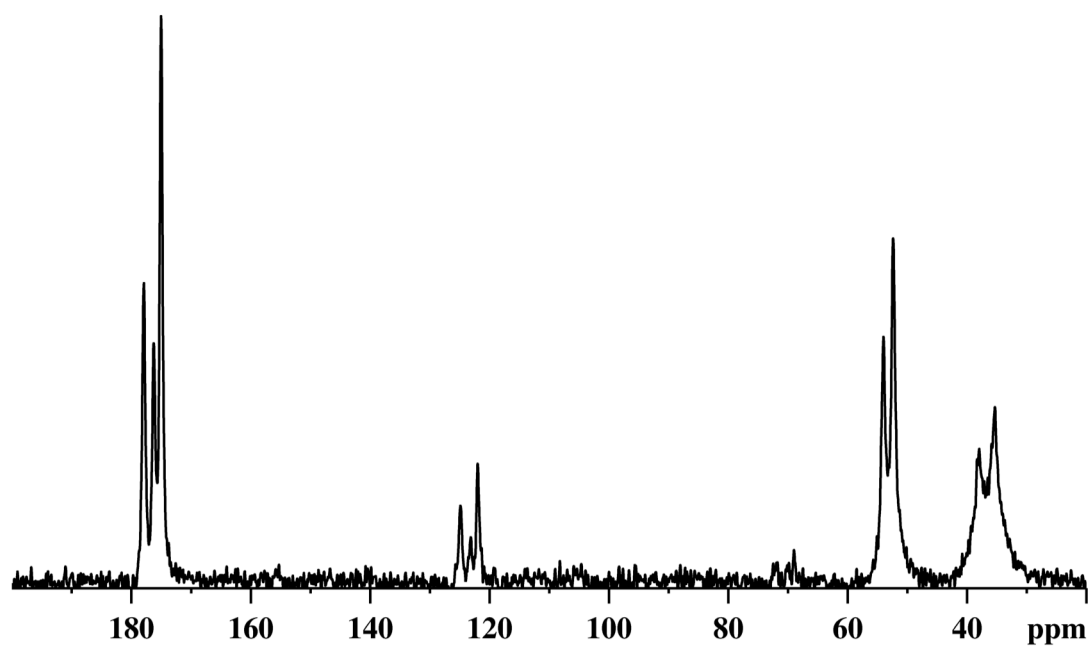


Figure 6.7 CP-MAS ^{13}C NMR spectrum of crystals from a solution of DL-aspartic acid, rapidly evaporated at 91°C , showing presence of both racemate and conglomerate crystals.

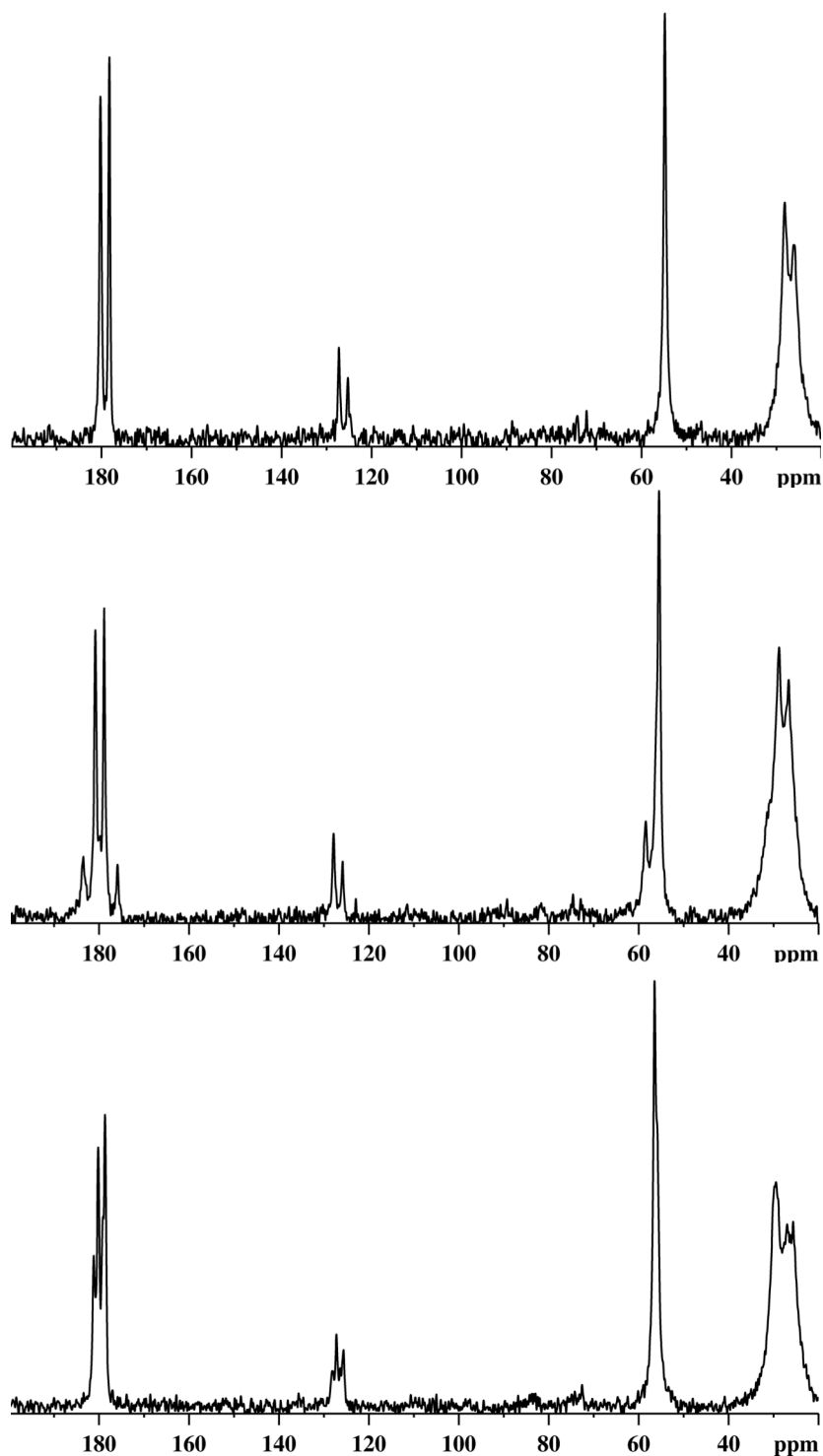


Figure 6.8 CP-MAS ^{13}C NMR spectra of (top) commercial L-glutamic acid (middle) DL-glutamic acid rapidly evaporated at 25 °C, showing presence of conglomerate and an as yet unidentified phase (bottom) DL-glutamic acid slowly evaporated at 25 °C, showing presence of conglomerate and another as yet unidentified phase.

References

- (1) H. D. Flack. Louis Pasteur's discovery of molecular chirality and spontaneous resolution in 1848, together with a complete review of his crystallographic and chemical work. *Acta Cryst.* **2009**, *A65*, 371–389.
- (2) W. A. Bonner. The origin and amplification of biomolecular chirality. *Origins of Life and Evolution of the Biosphere* **1991**, *21*, 59-111.
- (3) J. D. Dunitz, A. Gavezzotti. Proteogenic Amino Acids: Chiral and Racemic Crystal Packings and Stabilities. *J. Phys. Chem. B* **2012**, *116*, 6740–6750.
- (4) C. P. Brock, W. B. Schweizer, J. D. Dunitz. On the Validity of Wallach's Rule: On the Density and Stability of Racemic Crystals Compared with Their Chiral Counterparts. *J. Am. Chem. Soc.* **1991**, *113*, 9811-9820.
- (5) A. C. D. Newman, H. M. Powell. The spontaneous optical resolution of solvated tri-*o*-thymotide. *J. Chem. Soc.*, **1952**, 3747-3751.
- (6) T. Lee, Y. K. Lin. The Origin of Life and the Crystallization of Aspartic Acid in Water. *Crystal Growth & Design*, **2010**, *10*, 1652-1660.
- (7) A. C. F. Ribeiro, M. C. F. Barros, L. M. P. Verissimo, V. M. M. Lobo, A. J. M. Valente. Binary Diffusion Coefficients for Aqueous Solutions of L-Aspartic Acid and Its Respective Monosodium Salt. *J Solution Chem.* **2014**, *43* 83–92.
- (8) J. P. Amenda, H. C. Helgeson. Solubilities of the common L- α -amino acids as a function of temperature and solution pH. *Pure and Appl. Chem.* **1997**, *69*, 935-942.
- (9) M. S. Dunn, F. J. Ross, L. S. Read. The solubility of the amino acids in water. *J. Biol. Chem.* **1933**, *103*, 579-595.
- (10) A. Apelblata, E. Manzurola. Solubilities of L-aspartic, DL-aspartic, DL-glutamic, p-hydroxybenzoic, o-anistic, p-anistic, and itaconic acids in water from T = 278K to T = 345K *J. Chem. Thermodynamics*, **1997**, *29*, 1527-1533.
- (11) J. L. Derissen, H. J. Endeman, A. F. Peerdeman. The crystal and molecular structure of L-aspartic acid, *Acta Crystallographica* **1968**, *B24*, 1349-1353.
- (12) E. Bendeif, C. Jelsch, The experimental library multipolar atom model refinement of L-aspartic acid, *Acta Cryst.* **2007**, *C63*, o361-o364.
- (13) S.T. Rao. Refinement of DL-Aspartic Acid. *Acta Cryst.* **1973**, *B29*, 1718.
- (14) R. Flaig, T. Koritsanszky, D. Zobel, P. Luger. Topological Analysis of the Experimental Electron Densities of Amino Acids. 1. D,L-Aspartic Acid at 20 K. *J. Am. Chem. Soc.* **1998**, *120*, 2227-2238.
- (15) A. Sequeira, H. Rajagopal, M. Ramanadham. A Neutron Study on the Structure of DL-Aspartic Acid *Acta Cryst.* **1989**, *C45*, 906-908.
- (16) G.-M. Wang, Z.-X. Li, C.-S. Duan, H. Li. DL-Aspartic acid. *Acta Cryst.* **2007**, *E63*, o4003.

-
- (17) C. Sano. History of glutamate production. *Am. J. Clin. Nutr.* **2009**, *90*, 728S-732S.
- (18) J. D. Bernal. The crystal structure of the natural amino acids and related compounds. *Z Kristallogr.* **1931**, *78*, 363-369.
- (19) M. S. Lehmann, A. C. Nunes. A short hydrogen bond between near identical carboxyl groups in the α -modification of L-glutamic acid *Acta Cryst.* **1980**, *B36*, 1621-1625.
- (20) S. Hirokawa. A new modification of L-glutamic acid and its crystal structure. *Acta Cryst.* **1955**, *8*, 637-641.
- (21) J. D. Dunitz, W. B. Schweizer. Anhydrous DL-Glutamic Acid. *Acta Cryst.* **1995**, *C51*, 1377-1379.
- (22) C. Guo, J. Persons, G. S. Harbison. Helical chirality in hexamethylene triperoxide diamine. *Magn. Reson. Chem.* **2006**, *44*, 832 – 837.
- (23) L. E. Díaz, F. Morin, C. L. Mayne, D. M. Grant, C.-J. Chang. Conformational analysis of DL-, L- and D-methionine by solid-state ^{13}C NMR spectroscopy. *Magn. Reson. Chem.* **1986**, *24*, 167-170.
- (24) C.-C. Chen, Y. Zhu, L. B. Evans. Phase Partitioning of Biomolecules: Solubilities of Amino Acids. *Biotech. Progr.* **1989**, *5*, 111-118.
- (25) Z. T. Gu, K. Ebisawa, A. McDermott. Hydrogen bonding effects on amine rotation rates in crystalline amino acids. *Solid State Nuclear Magnetic Resonance* **1996**, *7*, 161-172.